NOAA FISHERIES

Electronic Reporting in Survey Research Applied to Estimating Fishing Effort

Marine Recreational Information Program

FY-2016

Project: Electronic Reporting in Survey Research Applied to Estimating Fishing Effort

J. Michael Brick - Author, Westat

Table of Contents

1. Is it Influential Scientific Information?

2. Has it had sufficient Peer Review?

3. Report Title

4. Executive Summary

5. Background

6. Methods

7. Results

8. Discussion/Conclusions/Recommendations

9. References

10. Appendix

Electronic Reporting in Survey Research Applied to Estimating Fishing Effort

1. Is it Influential Scientific Information?

N

2. Has it had sufficient Peer Review?

N

3. Report Title

Electronic Reporting in Survey Research Applied to Estimating Fishing Effort

4. Executive Summary

For the full report, please refer to the attached appendix.

5. Background

For the full report, please refer to the attached appendix.

6. Methods

7. Results

8. Discussion/Conclusions/Recommendations

For the full report, please see the attached appendix.

9. References

For the full report, please refer to the attached appendix.

10. Appendix

"Electronic Reporting in Survey Research Applied to Estimating Fishing Effort", page 1

Electronic Reporting in Survey Research Applied to Estimating Fishing Effort

J. Michael Brick

Westat

August 2, 2018

1. Introduction

Changes in the survey environment have resulted in lower survey response rates at the same time expectations have increased that newer modes of communication should improve the quality and convenience for respondents. This convergence has fueled a great deal of survey research over the past few years.

The National Oceanic and Atmospheric Administration's National Marine Fisheries Service (NMFS) is responsible for producing high quality and timely data for assessment and management of marine fish stocks. The NMFS' Marine Recreational Information Program (MRIP) collects information on marine recreational angling. In its 2006 report, the National Research Council (2006) suggested major changes in the random-digit-dial telephone methodology that was used to estimate fishing effort. The recommendations included, amongst others, using data from angler registries for sampling purposes and research on panels to collect data from the same sampled units over time. The National Academies (2017) reviewed the program and the methodological innovations in the MRIP about a decade later and found substantial progress had been made since 2006. This new review continued the call for more research into electronic data collection, noting "electronic data collection should be further evaluated as an option for the Fishing Effort Survey, including smartphone apps, electronic diaries for prospective data collection, and a web option for all or just panel members."

1.1 Estimating Fishing Effort

Fishing effort is the total number of recreational saltwater fishing trips taken during a specified time period. This total is the product of total number of persons in the population who took one or more recreational saltwater fishing trips during the period and the mean number of trips that they took during the period. Since the total population is known from Census Bureau data, this computation requires estimating prevalence (the percent of the population that fish) and the mean number of trips of those who did fish. These two components – prevalence and mean number of trips – are the target quantities in the design of off-site fishing effort surveys.

When the original National Research Council review was undertaken in 2006, a landline telephone survey was used to make these estimates. The report identified the low response rates and coverage rates of the telephone survey as being problematic. An additional criticism of the telephone survey in 2006 was that it did not take advantage of auxiliary data on fishing such as state fishing license data files.

The MRIP conducted a series of experiments testing alternative methods of collecting effort data in the years following the review. Those studies showed that a mail survey had much better response and coverage rates than the telephone survey. They also devised a method to use data from state fishing license files to enhance the precision of the estimates. The survey methodology that emerged from this research program is called the Fishing Effort Survey (FES). After a transition period, the FES became the standard fishing effort survey of MRIP (for estimating the numbers of shore and private boat fishing trips made by anglers) beginning in 2018.

1.2 Structure of Report

This review discusses two options that MRIP could explore to address recent advances in data collection methods. The first option looks at electronic data collection using non-probability sampling approaches. Non-probability methods have increased in interest primarily due to the low cost of collecting data over the Internet from large numbers of persons. Traditionally, non-probability samples have not been acceptable for most federal government surveys[1]. The second option is to use probability sampling methods, but to use different modes and designs that will allow greater use of electronic reporting by respondents.

Our review separately examines both the non-probability and the probability options. The actual data collection methods for both options greatly overlap since both make extensive use of the Internet. However, the two options are distinct in how they use the data to make inferences to the population under study. The non-probability option must rely on model assumptions to make inferences, while the probability sampling option uses the sample design, adjusted by model assumptions to handle missing data, as the basis for inferences. Hybrid designs (Fahimi 2015) that attempt to combine small probability samples and larger non-probability samples have been suggested, but there is very little research that can be used to discuss hybrid designs. The next section gives some background that is the basis for the following sections.

2. Background

2.1 Probability Samples

The standard paradigm for producing high quality estimates of finite population statistics starts with selecting a probability sample from a well-defined sampling frame, and then uses design-based inference methods based on these selection probabilities. This approach is built on the framework originally described by Neyman (1934) and has been used for the production of official statistics in the United States and almost all countries for decades. The probability sampling paradigm has an accepted theoretical foundation that has been expanded upon over the years to deal with imperfections such as incomplete coverage and nonresponse. These weighting adjustments or 'fixes'

[1] OMB Statistical Standard 1.2 states that "Any use of nonprobability sampling methods (e.g., cut-off or model-based samples) must be justified statistically and be able to measure estimation error."

to deal with missing data due to nonresponse and incomplete coverage are model-based (in that they assume a model such as the response propensity within a weighting class is constant). Generally, the assumption boils down to assuming the data are missing at random (MAR). For example, a common nonresponse adjustment involves creating weighting classes and assuming all the units in the weighting class either have the same response rate (response homogeneity) or have the same distribution of the outcome variables. If either of these assumptions holds, then the estimates are unbiased (Särndal, Swensson, and Wretman, 1992) when the sample size is large. Even with modest deviations from these assumptions, the adjustments remove much of the bias in many applications (Särndal and Lundström, 2005).

The probability sampling paradigm has been very successful, at least in part due to attributes that make it especially applicable for official statistics. One of the most powerful attractions of this approach is that the sampling and inference procedures are objective, thus avoiding many of the opportunities to inject biases based on the views of the researchers conducting the survey. Even with departures from the idealized structure due to nonresponse and coverage errors, standard procedures have been implemented in most probability samples to address these deficiencies that are relatively objective (Brick 2011).

Probability samples do not have to employ the same data collection methods that have been used for decades, primarily mail, telephone and face-to-face interviews. In fact, probability samples have embraced new technologies over time, especially the computerization of the systems of data collection. Research in moving more of the data collection effort to the Internet has been very active in the last 10 years or so (Tourangeau et al. 2017).

An example includes commercialized panels of respondents (commercialized means that the vendor sells access to the panel for a fee) who respond to surveys by the web. The first in the U.S. to do this was the KnowledgePanel of GfK, which has been around for almost 20 years. More recently, NORC introduced a panel called Amerispeak. Both of these panels recruit a probability sample of household members and then ask them to respond to a variety of surveys over time using the Web. When an organization wishes to do a survey, the organization pays the vendor to request some or all of the panel members be sent the survey to complete. Thus, the panel members may respond to multiple surveys developed for different organizations during their tenure in the panel.

Another example is the use of mobile apps (software deigned to on a smartphone) or Web apps (software designed to run on Web browsers) to collect data from respondents in probability samples. The FoodAPS project is an example of a project for the Department of Agriculture exploring the use of Food Log, a web-based tool that can be accessed by a Web browser or an app (the app is only available for a tablet and smartphone device). In this project, interviewers visit sampled household and do screening, an initial interview, and train respondents for responding to a diary of food consumption. Household members can request a loan of a smartphone and/or a barcode scanner (to scan food-packaging barcodes). A sampled household records all foods acquired for one week. At the end of the data collection week, interviewers go back to the sampled household, conduct a final interview, collect any loaned equipment, and give out incentives. Daily text/email reminders are sent to sampled members thanking those who participate and informing them of the amount of incentive accumulated. For those who didn't complete the daily Food Log, the messages reminded them to complete the Food Log to keep earning incentives.

The Bureau of Labor Statistics (BLS) is exploring a similar approach for replacing its paper diary that collects data on consumer expenditures. BLS decided that an app was not appropriate due to privacy and confidentiality concerns but has a Web app (that runs on a browser) diary that closely resembles an app when used on a mobile device (but does not require respondents to download an app).

2.2 Non-probability Samples

Although probability samples have been used in most official statistics, other strategies that do not rely on probabilities of selection have been used in many commercial and in some official establishment surveys (Stephan and McCarthy, 1958; Knaub, 2007). Collectively, such designs are often designated as non-probability samples (Baker et al., 2013). Non-probability samples have become increasingly common in recent years, as organizations have taken advantage of the ability to attract people to respond by using opt-in volunteer sampling methods. Baker et al. (2010) and Callegaro et al. (2014, Chapter 1) describe techniques used to intercept and enroll people to respond to surveys for opt-in Internet designs. The key and novel feature of internet non-probability samples is the ability to obtain large numbers of respondents very quickly and inexpensively.

For example, the cost to complete an interview of say 5,000 respondents from an existing vendor of non-probability samples on the Internet may be roughly 1 percent of the cost of a telephone survey and well less than 0.5 percent of a face-to-face survey. These cost estimates vary greatly depending on the vendor, the type of survey, and many other factors. Very little information on costs for these surveys is available in the literature.

Furthermore, the data collection often takes less than one week, while telephone and face-to-face interviews often stretch over months. This huge difference in costs and speed of collection is the reason non-probability sampling on the Internet has generated so much interest. More specifics about online, opt-in samples are presented in the next section.

Non-probability sampling theory is not as developed or consolidated, and many applications are largely a-theoretical. In these cases, the idea seems to be that a large sample size, regardless of how it is obtained, is sufficient to provide "good" estimates. The very low cost of the samples is what makes them attractive.

Researchers who have examined the properties of non-probability samples suggest selection bias is the most substantial contributor to errors (Bethlehem 2010). Some researchers have explored the theories of matching and estimation from observational studies to address selection bias. Rivers (2007) proposed matching in the selection of the non-probability respondents to the survey, like that used in epidemiological studies. Rivers sampled from a large, representative probability sample (an already completed government survey with standard sampling weights) and then selected Internet cases that matched the characteristics of the probability sample (usually demographic characteristics). This method is also related to quota sampling, where predesignated targets of the number of respondents are set up and data collection ends when these numbers have been reached. For example, the targets for age by sex groups might be used as the quotas.

In addition to demographic variables, some have explored other types of variables for matching and for weighting. Fahimi (2015) suggested using behavioral and attitudinal measures such as the use of coupons in shopping and the number of hours spent on the internet. Weighting methods such as poststratification, raking, and propensity score adjustments are sometimes used (Lee and Valiant 2009; Brick 2015). With a matched sample as suggested by Rivers (2007), the non-probability respondents could be assigned the weights from the probability sample (so the non-probability respondents are essentially substitute respondents to a probability sample). Mercer et al. (2017) explores the relationship between non-probability sampling and causal inference, and Elliott and Valliant (2017) describe weighting methods. Despite these types of investigations, practitioners have not accepted a standardized approach to sampling or estimation theory for non-probability samples (Buelens, Burger, and van den Brakel 2018).

Computing precision estimates from non-probability samples is also somewhat controversial, with some advocating no precision estimates should be provided while others have explored alternatives such as Bayesian credible intervals. As a result, many different variants exist and comparing them is difficult.

2.3 Online Non-probability Samples

Online non-probability surveys are very common, with sources suggesting that the majority of surveys, many of which are commercial market research surveys, are done online (Callegaro et al. 2014). Baker et al. (2010) provide a nice summary of the methods used to construct the samples for both one-time surveys and panels. They describe five activities:

- recruitment of members,
- joining procedures and profiling,
- specific study sampling,
- incentive programs, and
- panel maintenance.

The recruitment of members is a key component and is essentially the counterpoint to sample selection in a probability sample. The methods may be very selective (recruitment campaigns using advertising on specific websites or even using offline advertising) or more general depending on the target sample. Usually a contingent incentive or sweepstakes is used to attract persons. Some use a technique called co-registration agreements. Essentially, a website compiles emails of their visitors through a voluntary sign-up process and "monetize" this list by selling the emails. Others use "affiliate hubs" – sites that offer access to a number of different online merchants. Still another method relies on search engines, where a company buys ads that appear alongside search engine results hoping visitors will agree to do a survey. A less frequently used method is to request people they have already recruited to ask their friends or relatives to join, where they offer some award if the member adds others (this method is related to snowball sampling and respondent-driven sampling in some ways).

River sampling is sometimes considered a different approach, but only because it is done in real-time. Baker et al. (2010) define it as usually recruiting respondents when they are online, so it is related to the other methods using ads mentioned above. However, river sampling generally does not involve panel construction, but simply tries to encourage people to go to do some survey as they are surfing the Web. Those who do river sampling seldom have access to the demographics of the visitors, and so they rely on other companies to make the survey invitations.

Another option that straddles the border between a one-time and panel survey is an aggregator. This type of company works to share respondents from various sources (usually panel members) with those doing a survey. An aggregator can provide a list of email addresses that can be used to recruit respondents. While some of the characteristics of the members may be known along with the email address, the survey is essentially a one-time survey from the perspective of the researcher.

After the person has expressed an interest in taking a survey, a double opt-in process is almost always required to avoid surveys being done by computers. Double opt-in requires the person to sign up to do the survey and provide an email. The person then must take some positive action such as providing information sent to the email to get into the survey.

Once the person agrees and gets through the sign-in procedures, most surveys use some form of quota sampling to determine who should complete a specific survey. River sampling methods may route respondents to various surveys based on the (multiple) survey quotas. The quotas for many surveys are client specified and may involve demographics or other criteria, although most one-time surveys do not involve complicated quotas.

After recruitment, online probability panels (discussed in Section 4) and non-probability panels only differ materially in terms of follow-up procedures. Probability samples often use reminders and follow-ups to boost response rates; these methods may be expensive (incentives etc.) and take a bit more time. Non-probability surveys generally do not use follow-up methods and rely on having enough cases come through the door to complete the survey. The survey is done when the targeted number of respondents (meeting the quotas) has been satisfied. Often, the time required is a matter of a few days.

Most online, non-probability sampling vendors offer both one-time surveys and surveys from panels. The main difference between one-time non-probability samples and panels for most surveys is that panels usually capture profile data that allow them to subset to the appropriate subgroup needed for a particular survey. These profile data are sometimes used to match respondents as discussed in the previous section. For example, YouGov uses its profile data in this fashion for some of its applications.

While probability samples and panels usually use design-based weighting and variance estimation methods to produce estimates, many opt-in samples do not use weights at all. Brick (2015) describes a model-based survey framework that attempts to address weighting issues. He provides approaches for evaluating the assumptions underlying the models and associated weights. Other statistical models including likelihood and Bayesian methods (Wang et al. 2014) could be used, but examples under these models are rare.

The key strengths of online, non-probability sampling methods are its low-cost, very speedy data collection, and ability to obtain a large number of respondents. The measurement properties of the data collection are roughly equivalent to those from probability samples (online surveys have the advantage of avoiding interviewer effects and the ability to offer a wide range of visual displays).

Statistically, non-probability sampling has serious limitations that apply to both one-time samples and to panels. Baker et al. (2010) cautioned that the method should not be used for making population estimates ("Researchers should avoid non-probability online panels when one of the research objectives is to accurately estimate population values."). Bethlehem (2010) echoes this sentiment. Baker et al. (2013) are more nuanced in their discussion of the potential use of non-probability samples ("Non-probability samples may be appropriate for making statistical inferences, but the validity of the inferences rests on the appropriateness of the assumptions underlying the model and on how deviations from those assumptions affect the specific estimates.")

A major problem is that the methods used to recruit respondents for online non-probability samples are highly variable and subject to rapid change (Craig et al. 2013). Even using the "same" method, say placing ads on specific websites, will not necessarily result in the same diversity of respondents over time. The traffic to websites is highly dynamic. Aggregators chose different sources based on availability. Even things such as search engines are continuously being revised and these revisions may have consequences for making estimates, but they may not be obvious. For example, early work using the Google Search engine suggested it could be an early indicator of the severity of flu in the U.S., but this predictability evaporated when the internal mechanism of the search engine was modified unbeknownst to the researchers (Lazer et al. 2014).

2.3 Comparing Probability and Non-probability Samples

Several empirical studies have been conducted over the last 10 years or so to assess the performance of non-probability samples. Some early work (e.g., Yeager et al., 2011) compared probability and non-probability sample estimates to benchmarks and generally found the probability sampling estimates were more accurate. A more comprehensive and up-to-date review of comparisons is given by Callegaro et al. (2014, Chapter 2). Their review suggests that probability sampling, even with relatively low response rates, gives estimates that are closer to benchmark values than non-probability samples. These results are not consistent across all types of estimates and the differences may not be meaningful in some cases.

The comparisons for other statistics such as measures of association and trends are more ambiguous, partially because there are few reliable benchmarks for comparison. Most researchers have simply compared association or trend measures for non-probability samples to those from probability samples. Callegaro et al. (2014, Chapter 2) includes a review of the few published studies that look at measures of association. The results are less consistent than point estimates across the sampling methods. Some studies note that while there are differences, they are not large enough to influence policy decisions. For example, see Miller et al. (2010). For trends across time, there is scant methodological research evaluating the quality of trends. Some argue that theoretically trends and measures of association should not be very different whether estimated from probability or non-probability samples. This is clearly a conjecture and the conditions under which this holds have not been stated clearly. Nevertheless, the Centers for Disease Control have used non-probability online panels to assess trends for certain rare groups (e.g., Ding et al., 2015).

3. The Non-probability Option for Fishing Effort Surveys

This section begins with a discussion of online, non-probability sampling methods that could be used to survey anglers to estimate recreational fishing effort. We then discuss an extension of this type of design to the general concept of citizen science (similar to crowd-sourcing concepts) and the use of smartphone apps for collecting data. Mobile and Web apps are a data collection method rather than a sampling method, but the current literature does not clearly make that distinction.

3.1 Estimating Fishing Effort with Online Non-probability Samples

Selection bias is the major concern in most non-probability samples, and it would be a very serious issue in estimating fishing effort with an online non-probability sample. Research into models attempting to deal with selection bias thus far have been generally unsuccessful. Propensity models and propensity score weighting adjustments attempt to deal with selection bias by modeling those that have access to the Internet (Lee and Valliant, 2009) or those more likely to use the web heavily. Most of the research suggests selection bias is more complex than just a coverage issue, so only dealing with access to the Internet is insufficient. Furthermore, just being on the Internet often (as most respondents to non-probability surveys are) is not very predictive of being a respondent to non-probability sample recruitment.

Modeling participation in a survey without the type of active recruitment used in probability sampling is extremely difficult with few, if any, examples of this being done effectively. The problem is magnified for a fishing effort survey because angler surveys are more likely than many other types of surveys to suffer from avidity bias. We define avidity bias in this context as the overestimation of fishing prevalence that results when anglers are more likely to participate in the survey than non-anglers are. Groves et al. (2006) demonstrated the potential bias due to avidity in a birding survey experiment with a probability sample. In that example, people who were birders participated more than those who were not. In a non-probability sample without active recruitment and materials to encourage all people to respond, it is very likely that the selection bias would be a very serious problem.

A panel non-probability sample could be used, but the main benefit of a panel is the availability of profile data. Since both anglers and non-anglers are needed to estimate the percent of the population that fished, profile data that classified people as likely anglers is expected to have little value. Furthermore, the online panel respondents would be aware of the nature of the survey and avidity bias would be problematic. Respondents tend to be more willing to complete surveys on topics that they are familiar with or interest them.

The focus so far has been on selection bias due to avidity even though the effort estimate has both a prevalence component and a mean number of trips of those who did take trips. The earlier redesign efforts for the effort survey using probability sampling and different modes following the 2006 review showed that prevalence was much less stable than the mean number of trips of anglers. Prevalence was affected by the sampling frame, mode, and questionnaire while the mean number of trips per angler was not. As described above, we suspect this relationship may be exacerbated for non-probability samples, and the rationale for the avidity bias effect on prevalence has been provided. Our hypothesis is that the mean number of trips would be relatively robust for non-probability samples because it is conditional on having fished in the time period. This is just an hypothesis because no evaluation has been conducted with a non-probability sample.

An alternative approach that could be used to attempt to deal with avidity bias in a non-probability effort survey is to model the outcome variables of interest, given the set of respondents. There are two major reasons why this approach may be problematic. First, assuming the model holds over all people rather than just the respondents to the effort survey is tenuous; this assumption is not testable in most cases because we only observe the respondents. The respondents to a non-probability sample are not likely to be similar to the whole population, especially due to the avidity biases. Second, the modeling requirements for non-probability surveys can be extensive and rely heavily on good auxiliary information to be effective. Fishing register or license data are the only related source of such auxiliary information (demographic data typically are not very useful for modeling effort). While the fishing register data has been shown to be very valuable in sampling for the FES, it has serious limitations for use in modeling fishing prevalence. In particular, some states have very different rules and enforcement about who can fish and what the consequences are for recreational fishing without a license, so models would have to be very local. Such models would be subject to the same types of problems as the Google Flu example discussed earlier because local changes in enforcement or procedures for processing the data could have a big effect on the estimates.

Another major concern is the subjectivity that would be inherent in the modeling. The effort estimate has substantial consequences and interested parties could propose different model assumptions that result in estimates that more align with their interests. Since most of the modeling assumptions cannot be tested effectively, this could make it difficult to defend the estimates. Even if standard methods were developed for producing estimates from non-probability samples, there would be potential legitimate disagreements about those methods.

Producing an estimate without substantial bias is difficult with non-probability samples, but estimating the precision of the estimate is even more problematic. The usual, design-based estimate of variance is not appropriate because the sample is not selected with known probabilities. Without a measure of precision, it is impossible to assess whether policy decisions from the survey data are based on random error or the true measurement of fishing effort.

Another troubling issue is that bias, either selection bias or nonresponse bias, is likely to be a large component of the error. In this situation, estimating accuracy or mean square error is preferred to estimating variance. This estimation task is even more difficult because biases are so hard to estimate from all types of surveys. Unfortunately, most non-probability samples are not designed to give any estimate of bias or variance. If they report anything, it is usually just the variance computed as if a simple random sample had been selected. This approach is misleading.

A very different approach was suggested by Liu et al. (2017) that utilizes a non-probability sample and a probability sample to estimate catch (the method could be revised to estimate effort). Liu et al. (2017) describe this method in terms of a capture-recapture design. A large (non-probability) sample is 'captured' and report about their trips. A probability sample of anglers is conducted and the percentage of respondents in the probability sample who are 'recaptured' (were reporters in both samples) is estimated. Under some conditions and assumptions about response, unbiased estimates of the number of trips could be computed using both the non-probability and the probability sample.

The authors of the article consider the method within the context of estimating catch rather than effort, but the extension is feasible at least in theory. Several important conditions would be very hard to satisfy. One issue is the requirement for independence between the capture and recapture samples. It is not clear that the volunteers who respond (see next section on apps for more on a way to generate self-reporters) would be willing to respond to a second survey about their effort if sampled in the probability sample. It is also unclear whether the volunteer non-probability sample would be large enough to support this type of procedure. Another key issue is the method of matching the persons who responded to each survey. If this matching is not simple, the measurement error in matching could result in substantial bias. Liu et al. (2017) discuss ongoing evaluations of the method for catch that may provide some insights into these concerns, and we recommend waiting for results from these studies before trying to adapt this method for effort surveys.

3.2 Angler Apps

With the proliferation of smartphones and apps within the last decade, some researchers have been exploring using data from angler apps to provide estimates for marine recreational fishing activities. This approach to data collection often is categorized as citizen science, where citizens directly participate in various aspects of science. Citizen science covers a wide variety of projects engaging the general public or citizens in the practice of science. The types of activities range from participatory action research to large Web-enabled efforts. Examples include bird monitoring and the search for planets (Crain et al. 2014; Newman et al. 2012).

The use of apps for monitoring various forms of recreational fishing activities has been explored. Venturelli, Hyder and Skov (2017) discuss the utility of angler apps and the challenges associated with using them for monitoring activities. They define angler apps as mobile apps "that allow anglers to record, share and network" their activities. They provide thoughts on the status of this approach and suggest that creating standards and guidelines is important to take advantage of apps. The first challenge they identify is recruitment and retention of the citizen scientists to collect the data. For recruitment, they suggest making the apps "easy, fun and social." This approach is consistent with gamification that has been used to increase interest in many different fields, especially training (Hamari 2014). Keusch and Zhang (2017) examine the effects more specifically for online surveys and show that the benefits, while generally positive, are not very clear.

For producing effort estimates, any app also needs to appeal to those who do not fish recreationally since the key estimate is the percent who fished in the given time period. This issue is not discussed in any of the literature we reviewed and is a critical weakness in estimating fishing effort using this method. Venturelli, Hyder and Skov (2017) also suggest ways for improving retention of those who do use the app at least once. Here again the approach is likely to have serious disadvantages; the authors note that only 5 percent of those who begin to use the app still use it after 3 months. If even anglers do not persist in using the app its use in this context for estimating fishing effort is very dubious.

While the use of apps for other purposes may be reasonable and appropriate (Papenfuss et al. 2015; Jiorle, Ahrens and Allen 2016), our review shows this method is ill suited to the estimation of recreational fishing effort. Even if the improvements suggested by Venturelli, Hyder and Skov (2017) are eventually accomplished, it is extremely difficult to see how data from a fishing app will be useful to estimate the percent of the general population who take recreational saltwater fishing trips in a specified time period. This conclusion does not imply that fishing apps have no value for estimating fishing effort. In the next section, we discuss the possibility of using apps for data collection when respondents are recruited through a probability sampling approach.

4. Probability Approaches Using Electronic Modes

As noted previously, the empirical evidence shows that probability samples produce estimates that are closer to benchmark estimates than those from non-probability samples, although in some cases the differences are not that large. Because of the need to obtain responses from both anglers and non-anglers for effort surveys, avidity could easily bias both probability and non-probability samples. It is likely that the active recruitment and follow-up procedures used in probability samples, which are atypical in non-probability samples, gives a major advantage to probability sampling.

The current fishing effort survey uses probability sampling, but there are methods that could be researched within this paradigm that might enhance the survey. Some research into these methods are being investigated currently, but the field tests are planned for fall 2018 so no results are yet available. The applicability of some of the new methods depend on whether a cross-sectional design (one-time) or longitudinal design is used. We begin with cross-sectional designs and then discuss longitudinal designs.

4.1 Cross-Sectional Probability Designs for Fishing Effort

NMFS began research on a mail survey to collect data on fishing effort following the NRC report in 2006. The mail survey replaced the decades old random digit dial telephone survey of landline households, the Coastal Household Telephone Survey, completely in 2018. The evidence from the research shows the mail methodology has major response and coverage advantages over the telephone survey. Other mode changes, largely within the current cross-sectional, probability design, are being studied in an effort to further modernize and reduce data collection costs. In particular, MRIP is planning tests for using the Web to collect data.

Dillman (2017) summarizes his view of the future direction for using newer modes, primarily the Web, for data collection from cross-sectional probability samples. He also outlines some of the most important challenges. One approach is to use address-based sampling (ABS), as used in the mail FES, to sample households and then mail materials urging the respondents to go to the Web to complete the survey request. If this Web-push is followed by nonresponse attempts to have the household complete a mailed instrument, then it can be categorized as a mixed mode, Web-push survey.

Dillman (2017) talks largely about these mixed mode surveys, but he notes that mail instruments still result in the highest response rates. The MRIP research conducted following the 2006 also investigated mixed mode surveys, but MRIP chose the mail only approach because it had the highest response rates and coverage rates and generally had lower nonresponse bias than mixed mode methods. In particular, the option to offer respondents a choice of whether to respond by Web or by mail almost always depresses response rates (Medway and Fulton 2012) and was not deemed to be best practice for the fishing effort survey. Even though mail is still considered the best mode for data collection, Dillman (2017) highlights that changes in society are making mixed mode options more attractive. We start by discussing a Web-only design, and then move on to a Web-push design.

The Web-only design samples addresses from the ABS frame (derived from the United States Postal Service delivery sequence files) and mails materials to the sampled addresses urging the household to go online to complete a survey over the Web. If high response rates are desired, then multiple requests and monetary incentives are essential.

Web-only designs are often unable to achieve response rates that are sufficiently high for most sponsors. The lower than desired response rate is especially problematic because most studies show that Web-only survey respondents are not balanced leading to nonresponse bias in the estimates (Messer and Dillman 2011). The respondents tend to be younger and more highly educated than the general population.

Despite these concerns, several surveys have used a Web-only design, especially when funds for conducting the survey are limited. In 2016, Westat conducted a Web-only survey called the American National Election Survey (ANES) where funds were available to offer relatively large monetary incentives (both prepaid and promised) to boost response rates. The response rate for this 45-minute (on average) survey was 44 percent. The details on the methods and incentives are contained in the ANES methodology report[2]. The incentives for the ANES are very unlikely to be acceptable to OMB or MRIP for the administration of a fishing effort survey. The vast difference between the burden of the ANES and the FES would also argue against using such large incentives. However, the ANES experience does suggest that a Web-only design could achieve acceptable response rates and sample balance in terms of respondents if monetary incentives of sufficient size could be used. Most of the literature on experiments on incentives in surveys is for mail and telephone surveys, but that evidence shows that response rates increase at a lower rate as the amount of the incentive increases. The NMFS studied the effects of incentives for mail surveys and found that $1 and $2 pre-paid incentives raised response rates significantly and were cost-effective (lower overall cost per respondent compared to no incentive). See the appendix for a summary of these findings. Future research on the appropriate amount of prepaid and promised incentives for a Web-only or Web-push design is an area of research that could be informative.

Another feature of the ANES Web-only design was that the instrument was designed to allow respondents to use their smartphone. With a lengthy survey like the ANES, smartphones are often not the preferred mode of response, but since some respondents will only respond using a phone, this seems to be an essential design component. The fishing effort data requirements are much lower than the ANES, but still pose some challenges for smartphone response. As mentioned earlier, NMFS is exploring smartphone data collection options in the planned field test.

The mixed mode, Web-push design begins exactly like the Web-only design, but after one or more attempts to push the household to respond online, a mail questionnaire is sent to increase response rates. Many surveys are currently using this approach. The National Household Education Survey,

---

[2] www.electionstudies.org/studypages/anes_timeseries_2016/anes_timeseries_2016_methodology_report.pdf

which was a model in many ways for the FES in converting a telephone survey to a mail survey, tested this approach in 2016 (see its methodology report[3]).

The Web-push design is fast becoming the new default for cross-sectional surveys. It utilizes the Web to capture data inexpensively at the first phase and then tries to increase response and decrease bias by following up with mail. Generally speaking, the method does not suffer from differential mode effects because both the Web and mail are self-administered and visual modes. MRIP is planning to explore this option and is planning to field test in these areas.

4.2 Longitudinal Probability Designs

A longitudinal survey design opens new possibilities for data collection that are not as feasible with a one-time cross-sectional survey. The typical justification for most longitudinal surveys is the capability of producing estimates that cannot be estimated well from cross-sectional surveys, especially estimates of change. Some of the types of estimates of change of interest are spells (e.g., how long were people unemployed), estimates of gross change (e.g., how many people moved into and out of a job during a time period), and precise estimates of net change (e.g., by retaining sample members the estimates of change in employment status are more precise due to the positive correlation in the reported employment status over the time periods). The fishing effort survey is not required to produce these types of estimates, yet some potential advantages still exist for a longitudinal design. These potential advantages include:

- Obtaining a larger proportion of the respondent set who are likely to fish while still being able to produce unbiased estimates of the percent of the population that fished.

[3] nces.ed.gov/pubs2018/2018100.pdf

- Encouraging a larger proportion of the respondent set to respond by electronic means (Web or app).
- Reducing the cost of data collection.

These potential advantages could arise, but they are not guaranteed. Additional research is needed to determine the effectiveness of a longitudinal approach. Below, we outline the longitudinal design that we anticipate would be most likely to achieve some or all of these goals.

Most longitudinal surveys require substantial effort at both recruiting new sample and sample maintenance (keeping the sample members responding and avoiding attrition bias). A rotating panel design in which new members are recruited for each wave of data collection and some existing panel members are removed from the panel provides a balance between these requirements and is recommended for the research.

Figure 1 gives a picture of a rotating panel design. Each row represents a set of respondents coming into the panel (in this case the address is the sampled unit rather than the persons living at the household at a particular time) and a column represents a wave of data collection. In this diagram, each set of respondents (responding households at the sampled addresses) comes into the panel and remains for four waves of data collection and then exits. By the fourth wave, the panel has reached its steady-state with an incoming set and three retained sets of respondents. As described below, a subsample of respondents can be retained rather than retaining all sampled households for four waves.

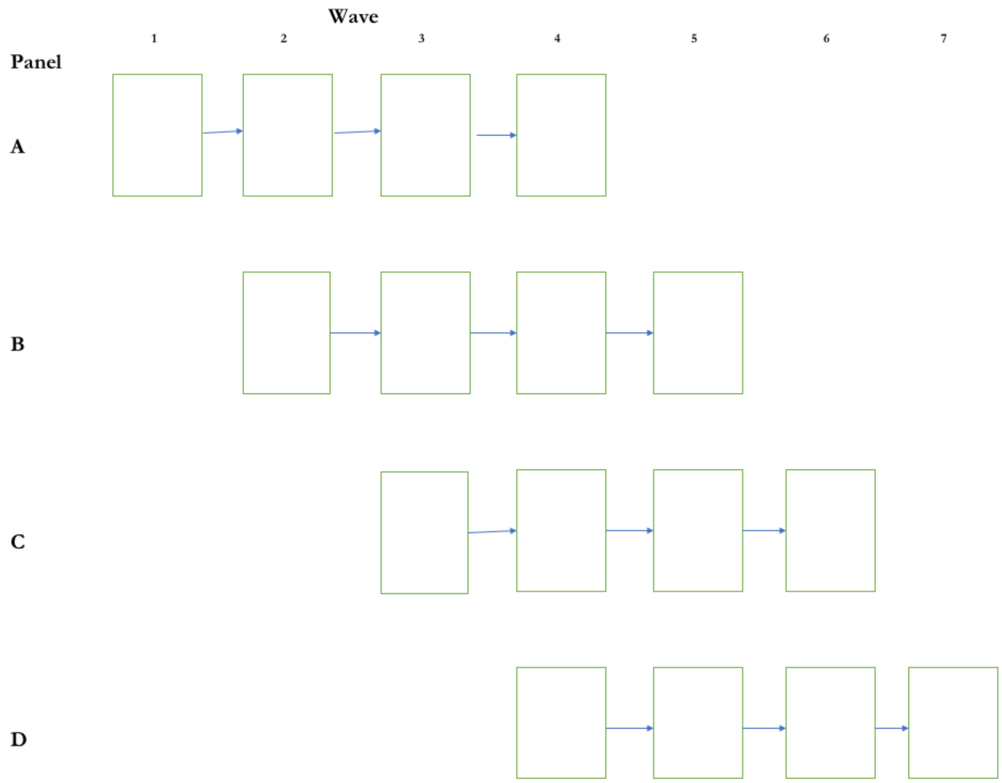

Figure 1. Rotating panel design schematic.

The key features of the particular rotating panel design suggested for the fishing effort survey are:

- In each time period, the sample would include an incoming or base wave of addresses to recruit new sample members, while also retaining a sample of respondents (addresses not people or households) from previous waves; respondents can be surveyed for a total of 4 waves (3 longitudinal follow-ups); all respondents who have completed 4 waves of data collection are dropped from the sample.
- The incoming wave is recruited using a Web-push design, and emails and text contact data are obtained from the household for subsequent rounds.
- A subsample of respondents from the incoming wave is retained for the subsequent three follow-ups; households with anglers and those who respond electronically can be all retained for follow-up and a subsample of the others are retained for follow-up. This both reduces costs and increase the sample size of anglers.
- Once the subsample is selected for the first follow-up, they are retained for the second and third follow-ups provided they respond; research on retaining subsampled households that fail to respond to the follow-ups in subsequent follow-ups should be conducted.
- The design can also be modified to include the households who are matched to the license registry data. These households are already much more likely to have anglers, so a possible scheme is to retain all of those households for the follow-ups regardless of fishing activity or mode of reporting. If this procedure is implemented, then a much smaller incoming sample is needed each wave to replace nonresponding and exiting panel members.

The effort data collection is relatively simple in terms of length and flow compared to many other surveys, although it has complexities in measurement such as recall of trips, the placement of the trips in time, and reporting for all adults in a household. Nevertheless, the survey is well-suited to electronic data collection and potentially the use of smartphones as a response mode. The Web-push design encourages sample members to respond electronically, and allows for the capture of contact information for those who respond by mail to the incoming wave so that they can be converted to electronic modes for the follow-ups.

Response rates to the incoming wave and retention of the subsampled households for the follow-ups is always an issue in longitudinal designs. A previous experiment with a mail longitudinal survey to estimate annual fishing participation rates suggests a reasonable response rate can be achieved with the rotating panel design (Andrews et al. 2016). Based on the experience with the ANES, we believe that an incentive program would greatly enhance the response rates to the study. We would encourage both a nominal prepaid incentive with the initial Web-push and a promised incentive for completing the survey. For example, a $2 incentive in the initial mailing and a promise of $5 for completing the survey (and each of the follow-ups) would probably both increase response rates by 10 percent or more and be cost effective. Lower response rates in the incoming and follow-up waves means larger numbers of addresses have to be sampled and surveyed and this increases costs. Research could be conducted to determine the best combination of pre- and post-paid incentives for the effort survey.

The earlier NMFS research with the mail longitudinal survey showed that initial or base wave respondents continued to respond to the follow-up surveys at roughly the same rate irrespective of whether they fished or not at the base wave. This result is very encouraging because it shows that the longitudinal design may help reduce avidity bias or at least not increase the bias. The wave estimate of the prevalence is a weighted combination of all the respondents to the particular wave, so it combines both initial respondents and the follow-up respondents. Of course, the follow-up respondents would be weighted to account for any subsampling for the follow-up that might be done.

An attempt was made to estimate the costs for the rotating design described above. We concluded that the data collection costs for the rotating panel design could be lower than for a repeated cross-sectional survey[4]. Clearly, these assumptions need to be examined empirically. The costs also depend on the subsampling rate for the follow-up. The precision of the estimates of effort is affected by the subsampling. A full statistical evaluation of these effects is needed to inform the design.

Furthermore, rotating panel surveys introduce other costs that are not part of the data collection costs per se. For example, the additional complexity of the design has implications for estimation procedures and the possibility of different types of measurement error (conditioning effects) that may require spending money that would not be spent in cross-sectional designs. Another cost that is not considered is the cost associated with changing from one design to another. When the effort survey transitioned from a telephone survey to a mail survey a series of activities were undertaken to aid in that transition including running the systems in parallel for a time and analytic studies to calibrate the estimates from the two designs.

4.3 Other Longitudinal Probability Designs Options

We also considered the use of an existing probability-based panel such as KnowledgePanel run by GfK or Amerispeak by NORC. These are general-purpose panels and do not appear to be suited for the needs of a fishing effort survey. For example, the Amerispeak panel currently has only about 10,000 members and would be too small for the state-specific needs of a fishing effort survey. The response rates for both panels are also much lower than NMFS might desire. Perhaps even more

[4] Note that the longitudinal survey used to model the costs was done completely by mail. The lack of experience using electronic data collection adds considerable ambiguity to these cost estimates.

importantly, if NMFS plans to continue to monitor fishing effort on a continuous basis, having a panel specifically designed for its own purpose is more appropriate and cost-effective.

A major advantage of having a panel specifically for fishing effort is that, assuming adequate numbers of people can be enlisted as panel members, some of the objectives of improving measurement identified in the National Academy studies can be explored beyond the reporting of data electronically. For example, the smartphone app could be tested in a survey environment where avidity bias does not severely limit the potential advantages of an app. The smartphone app has capabilities such as prompting the respondents each week to identify any eligible trips. The benefit of this would be to address recall bias. Another, albeit currently less feasible option, is that panel members who use other existing apps might be able to link directly to that data to capture data needed for the fishing survey. A longitudinal probability sample design that controls the potential bias due to nonresponse opens the door to making these improvements.

5. Summary

This review has examined methods for accomplishing some of the modernization goals that the National Academies (2017) encouraged in their study of MRIP's surveys of fishing effort. In particular, they called for more research into electronic data collection, including smartphones, electronic diaries, and a web option for all or just panel members. Our review separated the sampling for fishing effort surveys into probability and non-probability samples, with variations within these categories.

Non-probability samples can be characterized as data capture systems that amass responses without a selection scheme that give each unit in the population a known likelihood of selection. As a result, the design-based inference procedures are not valid with non-probability designs. Statistical models must be used to make inferences from non-probability samples.

Two types of non-probability samples that could be used to estimate fishing effort were examined in Section 3. The first is the online, opt-in non-probability sample that has been examined critically by many researchers (e.g., Baker et al. 2013). In most of these opt-in, online samples, selection bias is the major concern. Essentially, the lack of control of the process for recruiting respondents results in biases in the estimates because the respondents are not representative of the population; modeling has been generally unsuccessful in removing this bias.

We concluded that selection bias would be a very serious issue in estimating fishing effort with an online non-probability sample. Selection bias is more complex than coverage alone; being on the Internet often is not very predictive of being a respondent to non-probability sample recruitment. Selection bias might be further exacerbated for a fishing effort survey because surveys of this type already tend to suffer from avidity bias more than surveys of other topics. Using a non-probability panel is unlikely to reduce this bias because the type of profile data that these panels have available have little value for predicting fishing activity. An alternative or supplementary approach to deal with selection bias in a non-probability fishing effort survey is to posit statistical models, but this approach would face severe challenges. The modeling assumptions would rely on powerful auxiliary information, but these variables do not exist. Any models constructed would be difficult to test and highly subjective. For a government survey that has important policy implications, such subjectivity is not desirable.

The second type of non-probability sample uses data from angler apps to produce estimates. This non-probability sample would have even more challenges for producing fishing effort estimates because an app of this nature would have virtually no appeal to those who do not fish recreationally. This feature results in an extreme version of selection bias and would greatly overestimate the percent who fished in the given time period. It is a critical weakness in estimating fishing effort using this method. We do not recommend further consideration of this method at this time. However, since a mobile app is really a data collection mode rather than a sampling approach, apps do have potential when used within a probability sampling method as discussed below.

The current FES is a mail probability sample survey that replaced the decades old random digit dial probability telephone survey. This change occurred after research and experiments showed the mail methodology had major response rate and coverage advantages. Further work on modernizing the design has also begun by testing electronic data collection to a larger extent.

One approach is to continue to use address-based probability sampling but to mail materials urging the respondents to go to the Web to complete the survey request. A Web-only, where respondents can only reply by answering on the Web (either on a computer or smartphone), and a Web-push mixed mode survey, where nonrespondents to the Web-push can respond by mail, are being explored. Both of these designs maintain the probability sample but allow electronic data collection.

Another way to maintain a probability design but maximize electronic data collection is to move from the current cross-sectional survey with independent samples every wave to a longitudinal survey design. If respondents can be enrolled in a longitudinal design then it might be possible to encourage a large percentage of them to report using the Web, or perhaps even using a mobile app for the follow-up waves. If this is possible, then the costs of data collection may also be reduced. We proposed a rotating panel design as having the greatest potential and suggested some research and testing of this option. This research could help determine whether the advantages of a longitudinal design are substantial enough to offset the disruptive effect of change in a survey.

Overall, we believe that additional efforts to modernize and increase the use of electronic reporting is very worthy of research and field tests. The probability sample designs, even with the lower response rates that have been observed over time, have major advantages over non-probability designs for fishing effort surveys. We would urge concentrating resources in probability sample designs that using the Web as a mode of reporting, and rotating panel designs that again have the potential to increase electronic reporting.

References

Andrews, R., Opsomer, J., Brick, J. M. Rusetti, J., Van Voorhees, D., and Aikin, R. 2016. Testing a new approach to recreational saltwater fishing participation estimation. Downloaded July 6 2018 https://www.st.nmfs.noaa.gov/pims/main/public?method=DOWNLOAD_FR_DATA&record_id=389

Baker, Reg, Stephen J. Blumberg, J. Michael Brick, Mick P. Couper, Melanie Courtright, J. Michael Dennis, Don Dillman et al. 2010. "Research synthesis AAPOR report on online panels." *Public Opinion Quarterly* 74, 711-781.

Baker, Reg, J. Michael Brick, Nancy A. Bates, Mike Battaglia, Mick P. Couper, Jill A. Dever, Krista J. Gile, and Roger Tourangeau. 2013. "Summary report of the AAPOR task force on non-probability sampling." *Journal of Survey Statistics and Methodology*, smt008.

Bethlehem, Jelke. 2010. "Selection bias in web surveys." *International Statistical Review,* 78, 161-188.

Brick J.Michael. 2011. "The future of survey sampling". *Public Opinion Quarterly*, 75, 872-88.

Brick, J.Michael. 2015. "Compositional model inference". *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pp. 299-307. Alexandria, VA: American Statistical Association.

Buelens, B., Burger, J. and van den Brakel, J.A. 2018. "Comparing inference methods for non-probability samples." *International Statistical Review,* 10.13094/SMIF-2018-00003.

Callegaro, Mario, Reg Baker, Jelke Bethlehem, Anja S. Göritz, Jon A. Krosnick, and Paul J. Lavrakas. 2014. "Online panel research." *Online panel research: A data quality perspective*, 1-22.

Craig, Benjamin M., Ron D. Hays, A. Simon Pickard, David Cella, Dennis A. Revicki, and Bryce B. Reeve. 2013. "Comparison of US panel vendors for online surveys." *Journal of Medical Internet Research* 15, 11.

Crain, R., Cooper, C. and Dickinson, J.L. 2014. "Citizen science: a tool for integrating studies of human and natural systems." *Annual Review of Environment and Resources*, *39*, 641-645.

Dillman, D.A. 2017. "The promise and challenge of pushing respondents to the Web in mixed-mode surveys." *Survey Methodology*,*43*, 3-30.

Ding, Helen, Carla L. Black, Sarah W. Ball, Sara M.A. Donahue, Rebecca Fink, Carolyn B. Bridges, Katherine E. Kahn, Peng-Jun Lu, Walter W. Williams, Denise J. Jamieson. (2015). "Pregnant women and flu vaccination, Internet Panel Survey, United States, November 2015." Downladed from www.cdc.gov/flu/fluvaxview/pregnant-women-nov2015.htm on December 18, 2015.

Elliott, M.R. and Valliant, R. 2017. "Inference for nonprobability samples." *Statistical Science*, *32*, 249-264.

Fahimi, Mansour. 2015. "Towards a new math for sample surveys in the 21st Century." Paper presented at American Association of Public Opinion Research conference.

Groves, R.M., Couper, M.P., Presser, S., Singer, E., Tourangeau, R., Acosta, G.P. and Nelson, L. 2006. "Experiments in producing nonresponse bias." *International Journal of Public Opinion Quarterly*, *70*, 720-736.

Hamari, J., Koivisto, J. and Sarsa, H. 2014. "Does gamification work?--a literature review of empirical studies on gamification." In *System Sciences (HICSS), 2014 47th Hawaii International*

Conference, downloaded from https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6758978 on June 26, 2018.

Jiorle, R.P., Ahrens, R.N. and Allen, M.S. 2016. "Assessing the utility of a smartphone app for recreational fishery catch data." *Fisheries*, *41*, 758-766.

Knaub, James R. 2007. "Model and Survey Performance Measurement by the RSE and RSESP." *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 2730-2736.

Keusch, F. and Zhang, C. 2017. "A review of issues in gamified surveys." *Social Science Computer Review*, *35*, 147-166.

Lazer D, Kennedy R, King G, Vespignani A. 2014. "The parable of Google Flu: traps in big data analysis." *Science*, 343, 1203-1205.

Lee, Sunghee, and Richard Valliant. 2009. "Estimation for volunteer panel web surveys using propensity score adjustment and calibration adjustment." *Sociological Methods & Research* 37, 319-343.

Liu, B., Stokes, L., Topping, T. and Stunz, G. 2017. "Estimation of a total from a population of unknown size and application to estimating recreational red snapper catch in Texas." *Journal of Survey Statistics and Methodology*, *5*, 350-371.

Medway, R.L. and Fulton, J. 2012. "When more gets you less: a meta-analysis of the effect of concurrent web options on mail survey response rates." *Public Opinion Quarterly*, *76*, 733-746.

Mercer, A.W., Kreuter, F., Keeter, S. and Stuart, E.A. 2017. "Theory and practice in nonprobability surveys: parallels between causal inference and survey inference." *Public Opinion Quarterly*, *81*, 250-271.

Messer, B.L. and Dillman, D.A. 2011. "Surveying the general public over the internet using address-based sampling and mail contact procedures." *Public Opinion Quarterly*, *75*, 429-457.

Miller, Peter G., Jennifer Johnston, Matthew Dunn, Craig L. Fry, and Louisa Degenhardt. (2010). "Comparing probability and non-probability sampling methods in ecstasy research: Implications for the Internet as a research tool." *Substance use & misuse* 45, 437-450.

The National Academies of Sciences, Engineering, and Medicine. 2017. *Review of the Marine Recreational Information Program*. Washington, DC: The National Academies Press. https://doi.org/10.17226/24640.

NRC (National Research Council). 2006. *Review of recreational fFisheries survey methods*. Washington, DC: National Academies Press.

Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S. and Crowston, K. 2012. "The future of citizen science: emerging technologies and shifting paradigms." *Frontiers in Ecology and the Environment*, *10*, 298-304.

Neyman, Jerzy. 1934. "On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection." *Journal of the Royal Statistical Society*, 558-625.

Papenfuss, J.T., Phelps, N., Fulton, D. and Venturelli, P.A. 2015. "Smartphones reveal angler behavior: a case study of a popular mobile fishing application in Alberta, Canada." *Fisheries*, *40*, pp.318-327.

Rivers, D. 2007. "Sampling for web surveys". White paper prepared from presentation given at the *2007 Joint Statistical Meetings*, Salt Lake City, UT. https://s3.amazonaws.com/yg-public/Scientific/Sample+Matching_JSM.pdf.

Särndal, Carl-Erik, and Sixten Lundström. *Estimation in surveys with nonresponse*. John Wiley & Sons, 2005.

Särndal, C. E., B. Swensson, and J. Wretman. 1992. *Model assisted survey sampling*. Springer, New York.

Stephan, Frederick F., and Philip J. McCarthy. 1958. "Sampling opinions: An analysis of survey procedure." John Wiley & Sons.

Tourangeau R, Maitland A, Rivero G, Sun H, Williams D, Yan T. 2017. "Web surveys by smartphone and tablets: Effects on survey responses." *Public Opinion Quarterly*, 81, 896-929.

Valliant, Richard. 2013. "Comment." *Journal of Survey Statistics and Methodology*, smt010.

Venturelli, P.A., Hyder, K. and Skov, C. 2017. "Angler apps as a source of recreational fisheries data: opportunities, challenges and proposed standards." *Fish and Fisheries*, *18*(3), pp.578-595.

Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. 2014. "Forecasting elections with non-representative polls." *International Journal of Forecasting*.

Yeager, David S., Jon A. Krosnick, LinChiat Chang, Harold S. Javitz, Matthew S. Levendusky, Alberto Simpser, and Rui Wang. 2011. "Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples." *Public Opinion Quarterly*, nfr020.

Appendix

Marine Recreational Information Program

Fishing Effort Survey

Experimental Testing

9/26/2013

The MRIP Fishing Effort Survey (MFES) was implemented in Massachusetts, New York, North Carolina and Florida in October, 2012 to test a revised data collection design for monitoring marine recreational fishing effort. The survey, which collects information for two-month reference waves, included two experiments during the first two study waves, wave 5 (Sept-Oct 2012) and wave 6 (Nov-Dec, 2012), to test different survey design features aimed at maximizing efficiency and minimizing nonresponse error. Specifically, the experiments tested two versions of the survey instrument and four levels of cash incentives. Details of the experiments are provided below.

Instrument Testing

The MFES included an experiment to test two versions of the survey instrument. The objective of the experiment was to identify the instrument that maximized overall response rates while minimizing the potential for nonresponse bias resulting from differential nonresponse between anglers and non-anglers. One version of the instrument (Saltwater Fishing Survey) utilized a “screen out” approach that quickly identifies anglers (and non-anglers) and encourages participation by minimizing the number of survey questions, particularly for non-anglers. Person-level information, including details about recent fishing activity and limited demographic information, is collected for all household residents, but only if someone in the household reported fishing during the reference wave. The second version (Weather and Outdoor Activity Survey) utilized an “engaging” approach that encourages response by broadening the scope of the questions to include both fishing and non-fishing questions. This version collects person-level information for all residents of sampled households, regardless of whether or not household residents participated in saltwater fishing. Each wave, sampled addresses were randomly assigned to one of the two questionnaire types, which were evaluated in terms of response rates and reported fishing activity.

Table 1 provides the weighted response rates (AAPOR RR1 after excluding undeliverable addresses) and estimated fishing prevalence (percentage of households with residents who reported fishing during the wave) for the two versions of the instrument. Overall, the Weather and Outdoor Activity Survey achieved a significantly higher response rate than the Saltwater Fishing Survey, and there was no significant difference between instruments in estimated prevalence. The lack of a significant difference between instruments for estimated prevalence suggests that the gain in response for the engaging instrument cannot be attributed to increased survey participation by either anglers or non-anglers, but that both groups are more likely to respond to the Weather and Outdoor Activity Survey than the Saltwater Fishing Survey.

We also compared response rates and prevalence between instruments both among and within subpopulations defined by whether or not sampled addresses could be matched to state databases of licensed saltwater anglers – subpopulations expected to distinguish between households with anglers and households with no anglers or less avid anglers. As expected, both response rates and estimated prevalence were higher in the matched subpopulation than the unmatched subpopulation, confirming that a population expected to be interested in the survey topic - households with licensed anglers - is more likely to respond to a fishing survey and report fishing activity than a population that excludes licensed anglers[5]. Because we can identify household license status prior to data collection, we can account for differential nonresponse between matched and unmatched households in the estimation design by treating matched an unmatched domains as strata (Lohr, 2009).

Table 1. Weighted response rates and estimated prevalence overall and by domain for two versions of the survey instrument.

	Saltwater Fishing Survey		Weather and Outdoor Activity Survey	
	(%)	(n)	(%)	(n)
Response Rate				
Overall	31.1 (0.4)	17,511	34.7 (0.4)*	17,510
Matched	45.4 (1.1)	3,160	45.0 (1.0)	3,247
Unmatched	30.3 (0.4)	14,351	34.0 (0.5)*	14,263
Prevalence				
Overall	13.4 (0.5)	5,943	14.1 (0.5)	6,498
Matched	49.9 (1.7)	1,491	48.5 (1.6)	1,552
Unmatched	11.2 (0.6)	4,452	12.2 (0.6)	4,946

Notes – (1) standard errors are in parentheses. (2) Domains are defined by matching ABS samples to state databases of licensed saltwater anglers.
*Significantly different from Saltwater Fishing Survey ($p<0.05$).

There were no significant differences between instruments for either response rate or prevalence within the matched domain, suggesting that the inclusion of non-fishing questions in the Weather and Outdoor Activity Survey did not have an impact on response by either anglers or non-anglers. In the unmatched domain, the response rate was significantly higher for the Weather and Outdoor Activity Survey than the Saltwater Fishing Survey. However, the higher response rate did not translate to lower or higher estimates of prevalence; estimates of prevalence were not significantly different between instruments within the domain. This suggests that the engaging instrument uniformly increased the probability of response for anglers and non-anglers within the unmatched domain.

Differential nonresponse to a survey request between subpopulations will result in nonresponse bias if the subpopulations are different with respect to the survey topic. In the MRIP Fishing Effort Survey, we account for differential nonresponse between matched and unmatched households during sampling – matched and unmatched subpopulations are treated as independent

[5] The classification of sample into domains is dependent upon matching ABS sample to license databases by address and telephone number. This process is unlikely to be 100% accurate, so the unmatched domain is likely to include some households with licensed anglers. The unmatched domain also includes households with residents who fish without a license.

strata. Subsequently, the potential for nonresponse bias is limited to differential nonresponse between anglers and non-anglers within the matched and unmatched subpopulations. While the Weather and Outdoor Activity Survey achieved a higher response rate than the Saltwater Fishing Survey, both overall and within the unmatched subpopulation, the gains in response do not appear to result from a higher propensity to respond to the survey by either anglers or non-anglers. As a result, we cannot conclude that one of the instruments is more or less likely to minimize differential nonresponse between anglers and non-anglers. However, higher response rates decrease the risk for nonresponse bias and either lower data collection costs (for a fixed sample size) or increase the precision of estimates (for a fixed cost)[6]. Consequently, we conclude that the Weather and Outdoor Activity Survey is superior to the Saltwater Fishing Survey and recommend that the instrument be utilized for subsequent survey waves. Because it collects person-level information for all residents of all sampled households, the Weather and Outdoor Activity Survey also supports post-stratification of survey weights to population controls, which is an additional benefit of this recommendation.

Incentive Testing

The MRIP Fishing Effort Survey included an experiment to test the impact of modest, prepaid cash incentives on survey response and survey measures. Each wave, sampled addresses were randomly allocated to incentive treatment groups of $1, $2, and $5, as well as a non-incentive control group. Incentives were only included in the initial survey mailing. As in the instrument experiment, the objective of the incentive testing was to identify an optimum level of incentive that maximizes overall response while controlling costs and minimizes the potential for nonresponse bias resulting from differential nonresponse between anglers and non-anglers. Response rates, estimated fishing prevalence and relative costs of completing an interview were compared among incentive treatments to quantify the impacts of incentives.

Table 2 shows weighted response rates and the results of a logistic regression model predicting the effects of incentives on the odds of obtaining a completed survey. Including an incentive in the initial survey mailing significantly increased the odds of receiving a completed survey, and the odds increased significantly as the incentive amount increased. Cash incentives of $1, $2, and $5 increased the odds of receiving a completed survey by 63%, 93% and 137%, respectively.

Table 2. Weighted response rates and odds of receiving a completed survey by incentive amount.

Incentive	Response Rate (%)	n	Odds Ratio	95 % CI
$0	22.6	8,760	1.00	
$1	32.2	8,737	1.63*	(1.51, 1.77)
$2	36	8,738	1.93*	(1.78, 2.09)
$5	40.8	8,786	2.37*	(2.18, 2.56)

*Significantly different from the $0 control ($p$<0.05). Results of pairwise comparisons are as follows: $1>$0 (p<0.05), $2>$1 (p<0.05), $5>$2 (p<0.05).

[6] Assuming that fixed costs are the same for the two instruments, which was the case in the experiment.

Previous studies (Groves et al., 2006) have demonstrated that prepaid cash incentives can motivate individuals with little or no interest in a survey topic to respond to a survey request. Subsequently, we hypothesized that incentives would have a larger impact on non-anglers than anglers, minimizing differential nonresponse between the two populations. We initially explored this hypothesis by comparing estimated fishing prevalence among incentive conditions, expecting that gains in response in the incentive conditions would translate to lower estimates of fishing prevalence. The results do not support this hypothesis; there were no significant differences in prevalence among incentive conditions (Table 3).

Table 3. Overall estimated fishing prevalence by incentive amount.

Incentive	Prevalence (%)	n
$0	12.8	2,154
$1	14.1	3,065
$2	13.6	3,415
$5	14.1	3,807

Note – Differences in prevalence among treatments are not significant (p=0.05)

We further explored the interaction of topic salience and incentives by examining response rates and estimated fishing prevalence for the incentive conditions within domains defined by whether or not sampled addresses could be matched to databases of licensed saltwater anglers. We expected incentives to have a more pronounced effect in the unmatched domain, a population less likely to have an interest in the survey topic, than in the matched domain. Table 4 shows that incentives increased the odds of receiving a completed survey in both the matched and unmatched subpopulations. However, the value of the incentive seems to be more important in the unmatched domain, where the odds of receiving a completed survey increased uniformly and significantly as the value of the incentive increased ($0<$1<$2<$5). In contrast, the incentive amount was less significant in the matched domain, where the odds of receiving a completed survey were relatively flat among incentive conditions. These results are consistent with our expectations and suggest that a population with a low propensity to respond to a fishing survey can be motivated to participate by cash incentives, and that the motivation may increase as the incentive amount increases.

Table 4. Odds of receiving a completed survey by level of incentive for sample that could and could not be matched to state databases of licensed anglers.

	Subpopulation	
Comparison Pair	Matched OR	Unmatched OR
$1 vs. $0	1.75**	1.63**
$2 vs. $0	2.01**	1.93**
$5 vs. $0	2.11**	2.39**
$2 vs. $1	1.15	1.18**
$5 vs. $1	1.21*	1.46**
$5 vs. $2	1.05	1.24**

Notes – The second value in the comparison pair is the reference value.
Significance: *p<0.05, **p<0.0001

As noted previously, we expected that the gains in response in the incentive conditions would translate to lower estimates of fishing prevalence, particularly in the unmatched subpopulation. Once again, the results are not consistent with expectations; differences in fishing prevalence among treatments were not significant in either the matched or unmatched domain (Table 5). The lack of an effect of incentives on fishing prevalence suggests that the gains in response associated with increasing incentive amounts are uniform between anglers and non-anglers.

Table 5. Estimated fishing prevalence by incentive amount for a population of anglers (matched) and non-anglers (unmatched).

	Subpopulation			
	Matched		Unmatched	
Incentive	(%)	(n)	(%)	(n)
$0	49.2	533	10.7	1,621
$1	50.3	779	12	2,286
$2	48.6	837	11.6	2,578
$5	48.2	894	12.4	2,913

Note – Within subpopulations differences in prevalence among treatments are not significant (*p*=0.05)

We also examined the effect of cash incentives on overall data collection costs, specifically the direct costs of printing, postage, and the cash incentives themselves. Table 6 shows that the $5 incentive provided the largest gain in response, but the gain came at a relative cost of approximately $0.15 per completed interview. In contrast, the additional costs of the $1 and $2 incentives (20% and 38% higher cost than the $0 control, respectively) are more than offset by the associated gains in the number of completed surveys (42% and 58%, respectively). In other words, including a $1 or $2 cash incentive in the initial survey mailing actually decreased the cost of receiving a completed survey by 22% and 20%, respectively. These cost savings, which are conservative[7], could be used to lower overall data collection costs (for a fixed sample size) or increase the precision of survey estimates (for a fixed cost).

Table 6. Effect of incentives on data collection costs

Incentive Amount	Relative Cost Difference	Relative Difference in Completed Surveys	Relative Cost per Completed Survey
$0	1.00	1	$1.00
$1	1.20	1.42	$0.78
$2	1.38	1.58	$0.80
$5	1.90	1.75	$1.15

Note – relative differences reflect the ratio of quantities (cost, completes) in the experimental treatments to the zero dollar control.

Including a modest prepaid cash incentive in survey mailings clearly has a positive effect on survey response rates; the odds of receiving a completed survey increased significantly as the incentive amount increased. We expected the incentives to have a greater effect on non-anglers than anglers and decrease the potential for nonresponse bias by minimizing differential nonresponse between these two populations. However, the results of the experiment suggest that incentives increase response propensities for non-anglers and anglers equally. While this result does not support our hypothesis, it does demonstrate that incentives can increase the quantity of data without having a negative impact on survey measures. The experiment also demonstrated that incentives can decrease overall data collection costs. Based upon these findings, we conclude that a $2 incentive is optimal in terms of both maximizing response rates and minimizing data collection costs.

[7] The cost comparison assumes that the non-incentive direct costs (postage and printing) are the same for all survey treatments and does not reflect the fact that incentive conditions may not require as many follow-up mailings.

References

Groves, R.M., M.P. Couper, S. Presser, E. Singer, R. Tourangeau, G.P. Piani, N. Lindsay. 2006. Experiments in producing nonresponse bias. Public Opinion Quarterly, 70: 720-736.

Lohr, S. 2009. Multiple Frame Surveys. Chapter 4 in Pfeffermann, D. (Ed.) *Handbook of Statistics: Sample Surveys Design, Methods and Applications* (vol. 29A). Elsevier, Amsterdam.

Electronic Reporting in Survey Research Applied to Estimating Fishing Effort

J. Michael Brick

Westat

August 2, 2018

1. Introduction

Changes in the survey environment have resulted in lower survey response rates at the same time expectations have increased that newer modes of communication should improve the quality and convenience for respondents. This convergence has fueled a great deal of survey research over the past few years.

The National Oceanic and Atmospheric Administration's National Marine Fisheries Service (NMFS) is responsible for producing high quality and timely data for assessment and management of marine fish stocks. The NMFS' Marine Recreational Information Program (MRIP) collects information on marine recreational angling. In its 2006 report, the National Research Council (2006) suggested major changes in the random-digit-dial telephone methodology that was used to estimate fishing effort. The recommendations included, amongst others, using data from angler registries for sampling purposes and research on panels to collect data from the same sampled units over time. The National Academies (2017) reviewed the program and the methodological innovations in the MRIP about a decade later and found substantial progress had been made since 2006. This new review continued the call for more research into electronic data collection, noting "electronic data collection should be further evaluated as an option for the Fishing Effort Survey, including smartphone apps, electronic diaries for prospective data collection, and a web option for all or just panel members."

1.1 Estimating Fishing Effort

Fishing effort is the total number of recreational saltwater fishing trips taken during a specified time period. This total is the product of total number of persons in the population who took one or more recreational saltwater fishing trips during the period and the mean number of trips that they took during the period. Since the total population is known from Census Bureau data, this computation requires estimating prevalence (the percent of the population that fish) and the mean number of trips of those who did fish. These two components – prevalence and mean number of trips – are the target quantities in the design of off-site fishing effort surveys.

When the original National Research Council review was undertaken in 2006, a landline telephone survey was used to make these estimates. The report identified the low response rates and coverage rates of the telephone survey as being problematic. An additional criticism of the telephone survey in 2006 was that it did not take advantage of auxiliary data on fishing such as state fishing license data files.

The MRIP conducted a series of experiments testing alternative methods of collecting effort data in the years following the review. Those studies showed that a mail survey had much better response and coverage rates than the telephone survey. They also devised a method to use data from state fishing license files to enhance the precision of the estimates. The survey methodology that emerged from this research program is called the Fishing Effort Survey (FES). After a transition period, the FES became the standard fishing effort survey of MRIP (for estimating the numbers of shore and private boat fishing trips made by anglers) beginning in 2018.

1.2 Structure of Report

This review discusses two options that MRIP could explore to address recent advances in data collection methods. The first option looks at electronic data collection using non-probability sampling approaches. Non-probability methods have increased in interest primarily due to the low cost of collecting data over the Internet from large numbers of persons. Traditionally, non-probability samples have not been acceptable for most federal government surveys[1]. The second option is to use probability sampling methods, but to use different modes and designs that will allow greater use of electronic reporting by respondents.

Our review separately examines both the non-probability and the probability options. The actual data collection methods for both options greatly overlap since both make extensive use of the Internet. However, the two options are distinct in how they use the data to make inferences to the population under study. The non-probability option must rely on model assumptions to make inferences, while the probability sampling option uses the sample design, adjusted by model assumptions to handle missing data, as the basis for inferences. Hybrid designs (Fahimi 2015) that attempt to combine small probability samples and larger non-probability samples have been suggested, but there is very little research that can be used to discuss hybrid designs. The next section gives some background that is the basis for the following sections.

2. Background

2.1 Probability Samples

The standard paradigm for producing high quality estimates of finite population statistics starts with selecting a probability sample from a well-defined sampling frame, and then uses design-based inference methods based on these selection probabilities. This approach is built on the framework originally described by Neyman (1934) and has been used for the production of official statistics in the United States and almost all countries for decades. The probability sampling paradigm has an accepted theoretical foundation that has been expanded upon over the years to deal with imperfections such as incomplete coverage and nonresponse. These weighting adjustments or 'fixes'

[1] OMB Statistical Standard 1.2 states that "Any use of nonprobability sampling methods (e.g., cut-off or model-based samples) must be justified statistically and be able to measure estimation error."

to deal with missing data due to nonresponse and incomplete coverage are model-based (in that they assume a model such as the response propensity within a weighting class is constant). Generally, the assumption boils down to assuming the data are missing at random (MAR). For example, a common nonresponse adjustment involves creating weighting classes and assuming all the units in the weighting class either have the same response rate (response homogeneity) or have the same distribution of the outcome variables. If either of these assumptions holds, then the estimates are unbiased (Särndal, Swensson, and Wretman, 1992) when the sample size is large. Even with modest deviations from these assumptions, the adjustments remove much of the bias in many applications (Särndal and Lundström, 2005).

The probability sampling paradigm has been very successful, at least in part due to attributes that make it especially applicable for official statistics. One of the most powerful attractions of this approach is that the sampling and inference procedures are objective, thus avoiding many of the opportunities to inject biases based on the views of the researchers conducting the survey. Even with departures from the idealized structure due to nonresponse and coverage errors, standard procedures have been implemented in most probability samples to address these deficiencies that are relatively objective (Brick 2011).

Probability samples do not have to employ the same data collection methods that have been used for decades, primarily mail, telephone and face-to-face interviews. In fact, probability samples have embraced new technologies over time, especially the computerization of the systems of data collection. Research in moving more of the data collection effort to the Internet has been very active in the last 10 years or so (Tourangeau et al. 2017).

An example includes commercialized panels of respondents (commercialized means that the vendor sells access to the panel for a fee) who respond to surveys by the web. The first in the U.S. to do this was the KnowledgePanel of GfK, which has been around for almost 20 years. More recently, NORC introduced a panel called Amerispeak. Both of these panels recruit a probability sample of household members and then ask them to respond to a variety of surveys over time using the Web. When an organization wishes to do a survey, the organization pays the vendor to request some or all of the panel members be sent the survey to complete. Thus, the panel members may respond to multiple surveys developed for different organizations during their tenure in the panel.

Another example is the use of mobile apps (software deigned to on a smartphone) or Web apps (software designed to run on Web browsers) to collect data from respondents in probability samples. The FoodAPS project is an example of a project for the Department of Agriculture exploring the use of Food Log, a web-based tool that can be accessed by a Web browser or an app (the app is only available for a tablet and smartphone device). In this project, interviewers visit sampled household and do screening, an initial interview, and train respondents for responding to a diary of food consumption. Household members can request a loan of a smartphone and/or a barcode scanner (to scan food-packaging barcodes). A sampled household records all foods acquired for one week. At the end of the data collection week, interviewers go back to the sampled household, conduct a final interview, collect any loaned equipment, and give out incentives. Daily text/email reminders are sent to sampled members thanking those who participate and informing them of the amount of incentive accumulated. For those who didn't complete the daily Food Log, the messages reminded them to complete the Food Log to keep earning incentives.

The Bureau of Labor Statistics (BLS) is exploring a similar approach for replacing its paper diary that collects data on consumer expenditures. BLS decided that an app was not appropriate due to privacy and confidentiality concerns but has a Web app (that runs on a browser) diary that closely resembles an app when used on a mobile device (but does not require respondents to download an app).

2.2 Non-probability Samples

Although probability samples have been used in most official statistics, other strategies that do not rely on probabilities of selection have been used in many commercial and in some official establishment surveys (Stephan and McCarthy, 1958; Knaub, 2007). Collectively, such designs are often designated as non-probability samples (Baker et al., 2013). Non-probability samples have become increasingly common in recent years, as organizations have taken advantage of the ability to attract people to respond by using opt-in volunteer sampling methods. Baker et al. (2010) and Callegaro et al. (2014, Chapter 1) describe techniques used to intercept and enroll people to respond to surveys for opt-in Internet designs. The key and novel feature of internet non-probability samples is the ability to obtain large numbers of respondents very quickly and inexpensively.

For example, the cost to complete an interview of say 5,000 respondents from an existing vendor of non-probability samples on the Internet may be roughly 1 percent of the cost of a telephone survey and well less than 0.5 percent of a face-to-face survey. These cost estimates vary greatly depending on the vendor, the type of survey, and many other factors. Very little information on costs for these surveys is available in the literature.

Furthermore, the data collection often takes less than one week, while telephone and face-to-face interviews often stretch over months. This huge difference in costs and speed of collection is the reason non-probability sampling on the Internet has generated so much interest. More specifics about online, opt-in samples are presented in the next section.

Non-probability sampling theory is not as developed or consolidated, and many applications are largely a-theoretical. In these cases, the idea seems to be that a large sample size, regardless of how it is obtained, is sufficient to provide "good" estimates. The very low cost of the samples is what makes them attractive.

Researchers who have examined the properties of non-probability samples suggest selection bias is the most substantial contributor to errors (Bethlehem 2010). Some researchers have explored the theories of matching and estimation from observational studies to address selection bias. Rivers (2007) proposed matching in the selection of the non-probability respondents to the survey, like that used in epidemiological studies. Rivers sampled from a large, representative probability sample (an already completed government survey with standard sampling weights) and then selected Internet cases that matched the characteristics of the probability sample (usually demographic characteristics). This method is also related to quota sampling, where predesignated targets of the number of respondents are set up and data collection ends when these numbers have been reached. For example, the targets for age by sex groups might be used as the quotas.

In addition to demographic variables, some have explored other types of variables for matching and for weighting. Fahimi (2015) suggested using behavioral and attitudinal measures such as the use of coupons in shopping and the number of hours spent on the internet. Weighting methods such as poststratification, raking, and propensity score adjustments are sometimes used (Lee and Valiant 2009; Brick 2015). With a matched sample as suggested by Rivers (2007), the non-probability respondents could be assigned the weights from the probability sample (so the non-probability respondents are essentially substitute respondents to a probability sample). Mercer et al. (2017) explores the relationship between non-probability sampling and causal inference, and Elliott and Valliant (2017) describe weighting methods. Despite these types of investigations, practitioners have not accepted a standardized approach to sampling or estimation theory for non-probability samples (Buelens, Burger, and van den Brakel 2018).

Computing precision estimates from non-probability samples is also somewhat controversial, with some advocating no precision estimates should be provided while others have explored alternatives such as Bayesian credible intervals. As a result, many different variants exist and comparing them is difficult.

2.3 Online Non-probability Samples

Online non-probability surveys are very common, with sources suggesting that the majority of surveys, many of which are commercial market research surveys, are done online (Callegaro et al. 2014). Baker et al. (2010) provide a nice summary of the methods used to construct the samples for both one-time surveys and panels. They describe five activities:

- recruitment of members,
- joining procedures and profiling,
- specific study sampling,
- incentive programs, and
- panel maintenance.

The recruitment of members is a key component and is essentially the counterpoint to sample selection in a probability sample. The methods may be very selective (recruitment campaigns using advertising on specific websites or even using offline advertising) or more general depending on the target sample. Usually a contingent incentive or sweepstakes is used to attract persons. Some use a technique called co-registration agreements. Essentially, a website compiles emails of their visitors through a voluntary sign-up process and "monetize" this list by selling the emails. Others use "affiliate hubs" – sites that offer access to a number of different online merchants. Still another method relies on search engines, where a company buys ads that appear alongside search engine results hoping visitors will agree to do a survey. A less frequently used method is to request people they have already recruited to ask their friends or relatives to join, where they offer some award if the member adds others (this method is related to snowball sampling and respondent-driven sampling in some ways).

River sampling is sometimes considered a different approach, but only because it is done in real-time. Baker et al. (2010) define it as usually recruiting respondents when they are online, so it is related to the other methods using ads mentioned above. However, river sampling generally does not involve panel construction, but simply tries to encourage people to go to do some survey as they are surfing the Web. Those who do river sampling seldom have access to the demographics of the visitors, and so they rely on other companies to make the survey invitations.

Another option that straddles the border between a one-time and panel survey is an aggregator. This type of company works to share respondents from various sources (usually panel members) with those doing a survey. An aggregator can provide a list of email addresses that can be used to recruit respondents. While some of the characteristics of the members may be known along with the email address, the survey is essentially a one-time survey from the perspective of the researcher.

After the person has expressed an interest in taking a survey, a double opt-in process is almost always required to avoid surveys being done by computers. Double opt-in requires the person to sign up to do the survey and provide an email. The person then must take some positive action such as providing information sent to the email to get into the survey.

Once the person agrees and gets through the sign-in procedures, most surveys use some form of quota sampling to determine who should complete a specific survey. River sampling methods may route respondents to various surveys based on the (multiple) survey quotas. The quotas for many surveys are client specified and may involve demographics or other criteria, although most one-time surveys do not involve complicated quotas.

After recruitment, online probability panels (discussed in Section 4) and non-probability panels only differ materially in terms of follow-up procedures. Probability samples often use reminders and follow-ups to boost response rates; these methods may be expensive (incentives etc.) and take a bit more time. Non-probability surveys generally do not use follow-up methods and rely on having enough cases come through the door to complete the survey. The survey is done when the targeted number of respondents (meeting the quotas) has been satisfied. Often, the time required is a matter of a few days.

Most online, non-probability sampling vendors offer both one-time surveys and surveys from panels. The main difference between one-time non-probability samples and panels for most surveys is that panels usually capture profile data that allow them to subset to the appropriate subgroup needed for a particular survey. These profile data are sometimes used to match respondents as discussed in the previous section. For example, YouGov uses its profile data in this fashion for some of its applications.

While probability samples and panels usually use design-based weighting and variance estimation methods to produce estimates, many opt-in samples do not use weights at all. Brick (2015) describes a model-based survey framework that attempts to address weighting issues. He provides approaches for evaluating the assumptions underlying the models and associated weights. Other statistical models including likelihood and Bayesian methods (Wang et al. 2014) could be used, but examples under these models are rare.

The key strengths of online, non-probability sampling methods are its low-cost, very speedy data collection, and ability to obtain a large number of respondents. The measurement properties of the data collection are roughly equivalent to those from probability samples (online surveys have the advantage of avoiding interviewer effects and the ability to offer a wide range of visual displays).

Statistically, non-probability sampling has serious limitations that apply to both one-time samples and to panels. Baker et al. (2010) cautioned that the method should not be used for making population estimates ("Researchers should avoid non-probability online panels when one of the research objectives is to accurately estimate population values."). Bethlehem (2010) echoes this sentiment. Baker et al. (2013) are more nuanced in their discussion of the potential use of non-probability samples ("Non-probability samples may be appropriate for making statistical inferences, but the validity of the inferences rests on the appropriateness of the assumptions underlying the model and on how deviations from those assumptions affect the specific estimates.")

A major problem is that the methods used to recruit respondents for online non-probability samples are highly variable and subject to rapid change (Craig et al. 2013). Even using the "same" method, say placing ads on specific websites, will not necessarily result in the same diversity of respondents over time. The traffic to websites is highly dynamic. Aggregators chose different sources based on availability. Even things such as search engines are continuously being revised and these revisions may have consequences for making estimates, but they may not be obvious. For example, early work using the Google Search engine suggested it could be an early indicator of the severity of flu in the U.S., but this predictability evaporated when the internal mechanism of the search engine was modified unbeknownst to the researchers (Lazer et al. 2014).

2.3 Comparing Probability and Non-probability Samples

Several empirical studies have been conducted over the last 10 years or so to assess the performance of non-probability samples. Some early work (e.g., Yeager et al., 2011) compared probability and non-probability sample estimates to benchmarks and generally found the probability sampling estimates were more accurate. A more comprehensive and up-to-date review of comparisons is given by Callegaro et al. (2014, Chapter 2). Their review suggests that probability sampling, even with relatively low response rates, gives estimates that are closer to benchmark values than non-probability samples. These results are not consistent across all types of estimates and the differences may not be meaningful in some cases.

The comparisons for other statistics such as measures of association and trends are more ambiguous, partially because there are few reliable benchmarks for comparison. Most researchers have simply compared association or trend measures for non-probability samples to those from probability samples. Callegaro et al. (2014, Chapter 2) includes a review of the few published studies that look at measures of association. The results are less consistent than point estimates across the sampling methods. Some studies note that while there are differences, they are not large enough to influence policy decisions. For example, see Miller et al. (2010). For trends across time, there is scant methodological research evaluating the quality of trends. Some argue that theoretically trends and measures of association should not be very different whether estimated from probability or non-probability samples. This is clearly a conjecture and the conditions under which this holds have not been stated clearly. Nevertheless, the Centers for Disease Control have used non-probability online panels to assess trends for certain rare groups (e.g., Ding et al., 2015).

3. The Non-probability Option for Fishing Effort Surveys

This section begins with a discussion of online, non-probability sampling methods that could be used to survey anglers to estimate recreational fishing effort. We then discuss an extension of this type of design to the general concept of citizen science (similar to crowd-sourcing concepts) and the use of smartphone apps for collecting data. Mobile and Web apps are a data collection method rather than a sampling method, but the current literature does not clearly make that distinction.

3.1 Estimating Fishing Effort with Online Non-probability Samples

Selection bias is the major concern in most non-probability samples, and it would be a very serious issue in estimating fishing effort with an online non-probability sample. Research into models attempting to deal with selection bias thus far have been generally unsuccessful. Propensity models and propensity score weighting adjustments attempt to deal with selection bias by modeling those that have access to the Internet (Lee and Valliant, 2009) or those more likely to use the web heavily. Most of the research suggests selection bias is more complex than just a coverage issue, so only dealing with access to the Internet is insufficient. Furthermore, just being on the Internet often (as most respondents to non-probability surveys are) is not very predictive of being a respondent to non-probability sample recruitment.

Modeling participation in a survey without the type of active recruitment used in probability sampling is extremely difficult with few, if any, examples of this being done effectively. The problem is magnified for a fishing effort survey because angler surveys are more likely than many other types of surveys to suffer from avidity bias. We define avidity bias in this context as the overestimation of fishing prevalence that results when anglers are more likely to participate in the survey than non-anglers are. Groves et al. (2006) demonstrated the potential bias due to avidity in a birding survey experiment with a probability sample. In that example, people who were birders participated more than those who were not. In a non-probability sample without active recruitment and materials to encourage all people to respond, it is very likely that the selection bias would be a very serious problem.

A panel non-probability sample could be used, but the main benefit of a panel is the availability of profile data. Since both anglers and non-anglers are needed to estimate the percent of the population that fished, profile data that classified people as likely anglers is expected to have little value. Furthermore, the online panel respondents would be aware of the nature of the survey and avidity bias would be problematic. Respondents tend to be more willing to complete surveys on topics that they are familiar with or interest them.

The focus so far has been on selection bias due to avidity even though the effort estimate has both a prevalence component and a mean number of trips of those who did take trips. The earlier redesign efforts for the effort survey using probability sampling and different modes following the 2006 review showed that prevalence was much less stable than the mean number of trips of anglers. Prevalence was affected by the sampling frame, mode, and questionnaire while the mean number of trips per angler was not. As described above, we suspect this relationship may be exacerbated for non-probability samples, and the rationale for the avidity bias effect on prevalence has been provided. Our hypothesis is that the mean number of trips would be relatively robust for non-probability samples because it is conditional on having fished in the time period. This is just an hypothesis because no evaluation has been conducted with a non-probability sample.

An alternative approach that could be used to attempt to deal with avidity bias in a non-probability effort survey is to model the outcome variables of interest, given the set of respondents. There are two major reasons why this approach may be problematic. First, assuming the model holds over all people rather than just the respondents to the effort survey is tenuous; this assumption is not testable in most cases because we only observe the respondents. The respondents to a non-probability sample are not likely to be similar to the whole population, especially due to the avidity biases. Second, the modeling requirements for non-probability surveys can be extensive and rely heavily on good auxiliary information to be effective. Fishing register or license data are the only related source of such auxiliary information (demographic data typically are not very useful for modeling effort). While the fishing register data has been shown to be very valuable in sampling for the FES, it has serious limitations for use in modeling fishing prevalence. In particular, some states have very different rules and enforcement about who can fish and what the consequences are for recreational fishing without a license, so models would have to be very local. Such models would be subject to the same types of problems as the Google Flu example discussed earlier because local changes in enforcement or procedures for processing the data could have a big effect on the estimates.

Another major concern is the subjectivity that would be inherent in the modeling. The effort estimate has substantial consequences and interested parties could propose different model assumptions that result in estimates that more align with their interests. Since most of the modeling assumptions cannot be tested effectively, this could make it difficult to defend the estimates. Even if standard methods were developed for producing estimates from non-probability samples, there would be potential legitimate disagreements about those methods.

Producing an estimate without substantial bias is difficult with non-probability samples, but estimating the precision of the estimate is even more problematic. The usual, design-based estimate of variance is not appropriate because the sample is not selected with known probabilities. Without a measure of precision, it is impossible to assess whether policy decisions from the survey data are based on random error or the true measurement of fishing effort.

Another troubling issue is that bias, either selection bias or nonresponse bias, is likely to be a large component of the error. In this situation, estimating accuracy or mean square error is preferred to estimating variance. This estimation task is even more difficult because biases are so hard to estimate from all types of surveys. Unfortunately, most non-probability samples are not designed to give any estimate of bias or variance. If they report anything, it is usually just the variance computed as if a simple random sample had been selected. This approach is misleading.

A very different approach was suggested by Liu et al. (2017) that utilizes a non-probability sample and a probability sample to estimate catch (the method could be revised to estimate effort). Liu et al. (2017) describe this method in terms of a capture-recapture design. A large (non-probability) sample is 'captured' and report about their trips. A probability sample of anglers is conducted and the percentage of respondents in the probability sample who are 'recaptured' (were reporters in both samples) is estimated. Under some conditions and assumptions about response, unbiased estimates of the number of trips could be computed using both the non-probability and the probability sample.

The authors of the article consider the method within the context of estimating catch rather than effort, but the extension is feasible at least in theory. Several important conditions would be very hard to satisfy. One issue is the requirement for independence between the capture and recapture samples. It is not clear that the volunteers who respond (see next section on apps for more on a way to generate self-reporters) would be willing to respond to a second survey about their effort if sampled in the probability sample. It is also unclear whether the volunteer non-probability sample would be large enough to support this type of procedure. Another key issue is the method of matching the persons who responded to each survey. If this matching is not simple, the measurement error in matching could result in substantial bias. Liu et al. (2017) discuss ongoing evaluations of the method for catch that may provide some insights into these concerns, and we recommend waiting for results from these studies before trying to adapt this method for effort surveys.

3.2 Angler Apps

With the proliferation of smartphones and apps within the last decade, some researchers have been exploring using data from angler apps to provide estimates for marine recreational fishing activities. This approach to data collection often is categorized as citizen science, where citizens directly participate in various aspects of science. Citizen science covers a wide variety of projects engaging the general public or citizens in the practice of science. The types of activities range from participatory action research to large Web-enabled efforts. Examples include bird monitoring and the search for planets (Crain et al. 2014; Newman et al. 2012).

The use of apps for monitoring various forms of recreational fishing activities has been explored. Venturelli, Hyder and Skov (2017) discuss the utility of angler apps and the challenges associated with using them for monitoring activities. They define angler apps as mobile apps "that allow anglers to record, share and network" their activities. They provide thoughts on the status of this approach and suggest that creating standards and guidelines is important to take advantage of apps. The first challenge they identify is recruitment and retention of the citizen scientists to collect the data. For recruitment, they suggest making the apps "easy, fun and social." This approach is consistent with gamification that has been used to increase interest in many different fields, especially training (Hamari 2014). Keusch and Zhang (2017) examine the effects more specifically for online surveys and show that the benefits, while generally positive, are not very clear.

For producing effort estimates, any app also needs to appeal to those who do not fish recreationally since the key estimate is the percent who fished in the given time period. This issue is not discussed in any of the literature we reviewed and is a critical weakness in estimating fishing effort using this method. Venturelli, Hyder and Skov (2017) also suggest ways for improving retention of those who do use the app at least once. Here again the approach is likely to have serious disadvantages; the authors note that only 5 percent of those who begin to use the app still use it after 3 months. If even anglers do not persist in using the app its use in this context for estimating fishing effort is very dubious.

While the use of apps for other purposes may be reasonable and appropriate (Papenfuss et al. 2015; Jiorle, Ahrens and Allen 2016), our review shows this method is ill suited to the estimation of recreational fishing effort. Even if the improvements suggested by Venturelli, Hyder and Skov (2017) are eventually accomplished, it is extremely difficult to see how data from a fishing app will be useful to estimate the percent of the general population who take recreational saltwater fishing trips in a specified time period. This conclusion does not imply that fishing apps have no value for estimating fishing effort. In the next section, we discuss the possibility of using apps for data collection when respondents are recruited through a probability sampling approach.

4. Probability Approaches Using Electronic Modes

As noted previously, the empirical evidence shows that probability samples produce estimates that are closer to benchmark estimates than those from non-probability samples, although in some cases the differences are not that large. Because of the need to obtain responses from both anglers and non-anglers for effort surveys, avidity could easily bias both probability and non-probability samples. It is likely that the active recruitment and follow-up procedures used in probability samples, which are atypical in non-probability samples, gives a major advantage to probability sampling.

The current fishing effort survey uses probability sampling, but there are methods that could be researched within this paradigm that might enhance the survey. Some research into these methods are being investigated currently, but the field tests are planned for fall 2018 so no results are yet available. The applicability of some of the new methods depend on whether a cross-sectional design (one-time) or longitudinal design is used. We begin with cross-sectional designs and then discuss longitudinal designs.

4.1 Cross-Sectional Probability Designs for Fishing Effort

NMFS began research on a mail survey to collect data on fishing effort following the NRC report in 2006. The mail survey replaced the decades old random digit dial telephone survey of landline households, the Coastal Household Telephone Survey, completely in 2018. The evidence from the research shows the mail methodology has major response and coverage advantages over the telephone survey. Other mode changes, largely within the current cross-sectional, probability design, are being studied in an effort to further modernize and reduce data collection costs. In particular, MRIP is planning tests for using the Web to collect data.

Dillman (2017) summarizes his view of the future direction for using newer modes, primarily the Web, for data collection from cross-sectional probability samples. He also outlines some of the most important challenges. One approach is to use address-based sampling (ABS), as used in the mail FES, to sample households and then mail materials urging the respondents to go to the Web to complete the survey request. If this Web-push is followed by nonresponse attempts to have the household complete a mailed instrument, then it can be categorized as a mixed mode, Web-push survey.

Dillman (2017) talks largely about these mixed mode surveys, but he notes that mail instruments still result in the highest response rates. The MRIP research conducted following the 2006 also investigated mixed mode surveys, but MRIP chose the mail only approach because it had the highest response rates and coverage rates and generally had lower nonresponse bias than mixed mode methods. In particular, the option to offer respondents a choice of whether to respond by Web or by mail almost always depresses response rates (Medway and Fulton 2012) and was not deemed to be best practice for the fishing effort survey. Even though mail is still considered the best mode for data collection, Dillman (2017) highlights that changes in society are making mixed mode options more attractive. We start by discussing a Web-only design, and then move on to a Web-push design.

The Web-only design samples addresses from the ABS frame (derived from the United States Postal Service delivery sequence files) and mails materials to the sampled addresses urging the household to go online to complete a survey over the Web. If high response rates are desired, then multiple requests and monetary incentives are essential.

Web-only designs are often unable to achieve response rates that are sufficiently high for most sponsors. The lower than desired response rate is especially problematic because most studies show that Web-only survey respondents are not balanced leading to nonresponse bias in the estimates (Messer and Dillman 2011). The respondents tend to be younger and more highly educated than the general population.

Despite these concerns, several surveys have used a Web-only design, especially when funds for conducting the survey are limited. In 2016, Westat conducted a Web-only survey called the American National Election Survey (ANES) where funds were available to offer relatively large monetary incentives (both prepaid and promised) to boost response rates. The response rate for this 45-minute (on average) survey was 44 percent. The details on the methods and incentives are contained in the ANES methodology report[2]. The incentives for the ANES are very unlikely to be acceptable to OMB or MRIP for the administration of a fishing effort survey. The vast difference between the burden of the ANES and the FES would also argue against using such large incentives. However, the ANES experience does suggest that a Web-only design could achieve acceptable response rates and sample balance in terms of respondents if monetary incentives of sufficient size could be used. Most of the literature on experiments on incentives in surveys is for mail and telephone surveys, but that evidence shows that response rates increase at a lower rate as the amount of the incentive increases. The NMFS studied the effects of incentives for mail surveys and found that $1 and $2 pre-paid incentives raised response rates significantly and were cost-effective (lower overall cost per respondent compared to no incentive). See the appendix for a summary of these findings. Future research on the appropriate amount of prepaid and promised incentives for a Web-only or Web-push design is an area of research that could be informative.

Another feature of the ANES Web-only design was that the instrument was designed to allow respondents to use their smartphone. With a lengthy survey like the ANES, smartphones are often not the preferred mode of response, but since some respondents will only respond using a phone, this seems to be an essential design component. The fishing effort data requirements are much lower than the ANES, but still pose some challenges for smartphone response. As mentioned earlier, NMFS is exploring smartphone data collection options in the planned field test.

The mixed mode, Web-push design begins exactly like the Web-only design, but after one or more attempts to push the household to respond online, a mail questionnaire is sent to increase response rates. Many surveys are currently using this approach. The National Household Education Survey,

---

[2] www.electionstudies.org/studypages/anes_timeseries_2016/anes_timeseries_2016_methodology_report.pdf

which was a model in many ways for the FES in converting a telephone survey to a mail survey, tested this approach in 2016 (see its methodology report[3]).

The Web-push design is fast becoming the new default for cross-sectional surveys. It utilizes the Web to capture data inexpensively at the first phase and then tries to increase response and decrease bias by following up with mail. Generally speaking, the method does not suffer from differential mode effects because both the Web and mail are self-administered and visual modes. MRIP is planning to explore this option and is planning to field test in these areas.

4.2 Longitudinal Probability Designs

A longitudinal survey design opens new possibilities for data collection that are not as feasible with a one-time cross-sectional survey. The typical justification for most longitudinal surveys is the capability of producing estimates that cannot be estimated well from cross-sectional surveys, especially estimates of change. Some of the types of estimates of change of interest are spells (e.g., how long were people unemployed), estimates of gross change (e.g., how many people moved into and out of a job during a time period), and precise estimates of net change (e.g., by retaining sample members the estimates of change in employment status are more precise due to the positive correlation in the reported employment status over the time periods). The fishing effort survey is not required to produce these types of estimates, yet some potential advantages still exist for a longitudinal design. These potential advantages include:

- Obtaining a larger proportion of the respondent set who are likely to fish while still being able to produce unbiased estimates of the percent of the population that fished.

[3] nces.ed.gov/pubs2018/2018100.pdf

- Encouraging a larger proportion of the respondent set to respond by electronic means (Web or app).
- Reducing the cost of data collection.

These potential advantages could arise, but they are not guaranteed. Additional research is needed to determine the effectiveness of a longitudinal approach. Below, we outline the longitudinal design that we anticipate would be most likely to achieve some or all of these goals.

Most longitudinal surveys require substantial effort at both recruiting new sample and sample maintenance (keeping the sample members responding and avoiding attrition bias). A rotating panel design in which new members are recruited for each wave of data collection and some existing panel members are removed from the panel provides a balance between these requirements and is recommended for the research.

Figure 1 gives a picture of a rotating panel design. Each row represents a set of respondents coming into the panel (in this case the address is the sampled unit rather than the persons living at the household at a particular time) and a column represents a wave of data collection. In this diagram, each set of respondents (responding households at the sampled addresses) comes into the panel and remains for four waves of data collection and then exits. By the fourth wave, the panel has reached its steady-state with an incoming set and three retained sets of respondents. As described below, a subsample of respondents can be retained rather than retaining all sampled households for four waves.

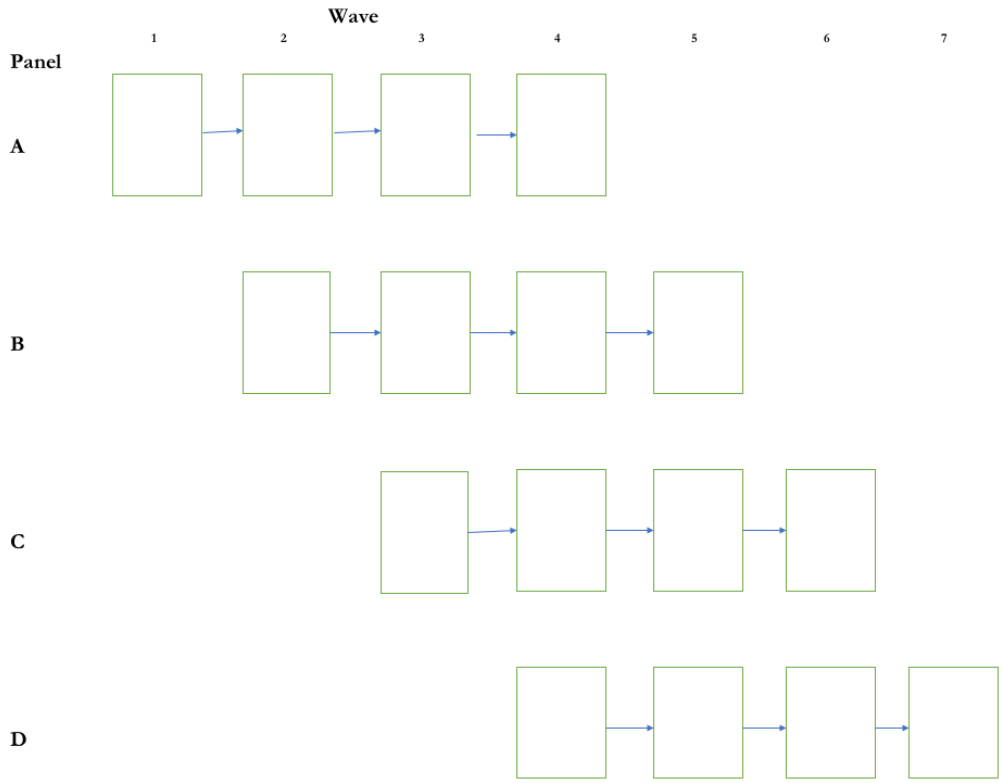

Figure 1. Rotating panel design schematic.

The key features of the particular rotating panel design suggested for the fishing effort survey are:

- In each time period, the sample would include an incoming or base wave of addresses to recruit new sample members, while also retaining a sample of respondents (addresses not people or households) from previous waves; respondents can be surveyed for a total of 4 waves (3 longitudinal follow-ups); all respondents who have completed 4 waves of data collection are dropped from the sample.
- The incoming wave is recruited using a Web-push design, and emails and text contact data are obtained from the household for subsequent rounds.
- A subsample of respondents from the incoming wave is retained for the subsequent three follow-ups; households with anglers and those who respond electronically can be all retained for follow-up and a subsample of the others are retained for follow-up. This both reduces costs and increase the sample size of anglers.
- Once the subsample is selected for the first follow-up, they are retained for the second and third follow-ups provided they respond; research on retaining subsampled households that fail to respond to the follow-ups in subsequent follow-ups should be conducted.
- The design can also be modified to include the households who are matched to the license registry data. These households are already much more likely to have anglers, so a possible scheme is to retain all of those households for the follow-ups regardless of fishing activity or mode of reporting. If this procedure is implemented, then a much smaller incoming sample is needed each wave to replace nonresponding and exiting panel members.

The effort data collection is relatively simple in terms of length and flow compared to many other surveys, although it has complexities in measurement such as recall of trips, the placement of the trips in time, and reporting for all adults in a household. Nevertheless, the survey is well-suited to electronic data collection and potentially the use of smartphones as a response mode. The Web-push design encourages sample members to respond electronically, and allows for the capture of contact information for those who respond by mail to the incoming wave so that they can be converted to electronic modes for the follow-ups.

Response rates to the incoming wave and retention of the subsampled households for the follow-ups is always an issue in longitudinal designs. A previous experiment with a mail longitudinal survey to estimate annual fishing participation rates suggests a reasonable response rate can be achieved with the rotating panel design (Andrews et al. 2016). Based on the experience with the ANES, we believe that an incentive program would greatly enhance the response rates to the study. We would encourage both a nominal prepaid incentive with the initial Web-push and a promised incentive for completing the survey. For example, a $2 incentive in the initial mailing and a promise of $5 for completing the survey (and each of the follow-ups) would probably both increase response rates by 10 percent or more and be cost effective. Lower response rates in the incoming and follow-up waves means larger numbers of addresses have to be sampled and surveyed and this increases costs. Research could be conducted to determine the best combination of pre- and post-paid incentives for the effort survey.

The earlier NMFS research with the mail longitudinal survey showed that initial or base wave respondents continued to respond to the follow-up surveys at roughly the same rate irrespective of whether they fished or not at the base wave. This result is very encouraging because it shows that the longitudinal design may help reduce avidity bias or at least not increase the bias. The wave estimate of the prevalence is a weighted combination of all the respondents to the particular wave, so it combines both initial respondents and the follow-up respondents. Of course, the follow-up respondents would be weighted to account for any subsampling for the follow-up that might be done.

An attempt was made to estimate the costs for the rotating design described above. We concluded that the data collection costs for the rotating panel design could be lower than for a repeated cross-sectional survey[4]. Clearly, these assumptions need to be examined empirically. The costs also depend on the subsampling rate for the follow-up. The precision of the estimates of effort is affected by the subsampling. A full statistical evaluation of these effects is needed to inform the design.

Furthermore, rotating panel surveys introduce other costs that are not part of the data collection costs per se. For example, the additional complexity of the design has implications for estimation procedures and the possibility of different types of measurement error (conditioning effects) that may require spending money that would not be spent in cross-sectional designs. Another cost that is not considered is the cost associated with changing from one design to another. When the effort survey transitioned from a telephone survey to a mail survey a series of activities were undertaken to aid in that transition including running the systems in parallel for a time and analytic studies to calibrate the estimates from the two designs.

4.3 Other Longitudinal Probability Designs Options

We also considered the use of an existing probability-based panel such as KnowledgePanel run by GfK or Amerispeak by NORC. These are general-purpose panels and do not appear to be suited for the needs of a fishing effort survey. For example, the Amerispeak panel currently has only about 10,000 members and would be too small for the state-specific needs of a fishing effort survey. The response rates for both panels are also much lower than NMFS might desire. Perhaps even more

[4] Note that the longitudinal survey used to model the costs was done completely by mail. The lack of experience using electronic data collection adds considerable ambiguity to these cost estimates.

importantly, if NMFS plans to continue to monitor fishing effort on a continuous basis, having a panel specifically designed for its own purpose is more appropriate and cost-effective.

A major advantage of having a panel specifically for fishing effort is that, assuming adequate numbers of people can be enlisted as panel members, some of the objectives of improving measurement identified in the National Academy studies can be explored beyond the reporting of data electronically. For example, the smartphone app could be tested in a survey environment where avidity bias does not severely limit the potential advantages of an app. The smartphone app has capabilities such as prompting the respondents each week to identify any eligible trips. The benefit of this would be to address recall bias. Another, albeit currently less feasible option, is that panel members who use other existing apps might be able to link directly to that data to capture data needed for the fishing survey. A longitudinal probability sample design that controls the potential bias due to nonresponse opens the door to making these improvements.

5. Summary

This review has examined methods for accomplishing some of the modernization goals that the National Academies (2017) encouraged in their study of MRIP's surveys of fishing effort. In particular, they called for more research into electronic data collection, including smartphones, electronic diaries, and a web option for all or just panel members. Our review separated the sampling for fishing effort surveys into probability and non-probability samples, with variations within these categories.

Non-probability samples can be characterized as data capture systems that amass responses without a selection scheme that give each unit in the population a known likelihood of selection. As a result, the design-based inference procedures are not valid with non-probability designs. Statistical models must be used to make inferences from non-probability samples.

Two types of non-probability samples that could be used to estimate fishing effort were examined in Section 3. The first is the online, opt-in non-probability sample that has been examined critically by many researchers (e.g., Baker et al. 2013). In most of these opt-in, online samples, selection bias is the major concern. Essentially, the lack of control of the process for recruiting respondents results in biases in the estimates because the respondents are not representative of the population; modeling has been generally unsuccessful in removing this bias.

We concluded that selection bias would be a very serious issue in estimating fishing effort with an online non-probability sample. Selection bias is more complex than coverage alone; being on the Internet often is not very predictive of being a respondent to non-probability sample recruitment. Selection bias might be further exacerbated for a fishing effort survey because surveys of this type already tend to suffer from avidity bias more than surveys of other topics. Using a non-probability panel is unlikely to reduce this bias because the type of profile data that these panels have available have little value for predicting fishing activity. An alternative or supplementary approach to deal with selection bias in a non-probability fishing effort survey is to posit statistical models, but this approach would face severe challenges. The modeling assumptions would rely on powerful auxiliary information, but these variables do not exist. Any models constructed would be difficult to test and highly subjective. For a government survey that has important policy implications, such subjectivity is not desirable.

The second type of non-probability sample uses data from angler apps to produce estimates. This non-probability sample would have even more challenges for producing fishing effort estimates because an app of this nature would have virtually no appeal to those who do not fish recreationally. This feature results in an extreme version of selection bias and would greatly overestimate the percent who fished in the given time period. It is a critical weakness in estimating fishing effort using this method. We do not recommend further consideration of this method at this time. However, since a mobile app is really a data collection mode rather than a sampling approach, apps do have potential when used within a probability sampling method as discussed below.

The current FES is a mail probability sample survey that replaced the decades old random digit dial probability telephone survey. This change occurred after research and experiments showed the mail methodology had major response rate and coverage advantages. Further work on modernizing the design has also begun by testing electronic data collection to a larger extent.

One approach is to continue to use address-based probability sampling but to mail materials urging the respondents to go to the Web to complete the survey request. A Web-only, where respondents can only reply by answering on the Web (either on a computer or smartphone), and a Web-push mixed mode survey, where nonrespondents to the Web-push can respond by mail, are being explored. Both of these designs maintain the probability sample but allow electronic data collection.

Another way to maintain a probability design but maximize electronic data collection is to move from the current cross-sectional survey with independent samples every wave to a longitudinal survey design. If respondents can be enrolled in a longitudinal design then it might be possible to encourage a large percentage of them to report using the Web, or perhaps even using a mobile app for the follow-up waves. If this is possible, then the costs of data collection may also be reduced. We proposed a rotating panel design as having the greatest potential and suggested some research and testing of this option. This research could help determine whether the advantages of a longitudinal design are substantial enough to offset the disruptive effect of change in a survey.

Overall, we believe that additional efforts to modernize and increase the use of electronic reporting is very worthy of research and field tests. The probability sample designs, even with the lower response rates that have been observed over time, have major advantages over non-probability designs for fishing effort surveys. We would urge concentrating resources in probability sample designs that using the Web as a mode of reporting, and rotating panel designs that again have the potential to increase electronic reporting.

References

Andrews, R., Opsomer, J., Brick, J. M. Rusetti, J., Van Voorhees, D., and Aikin, R. 2016. Testing a new approach to recreational saltwater fishing participation estimation. Downloaded July 6 2018 https://www.st.nmfs.noaa.gov/pims/main/public?method=DOWNLOAD_FR_DATA&record_id=389

Baker, Reg, Stephen J. Blumberg, J. Michael Brick, Mick P. Couper, Melanie Courtright, J. Michael Dennis, Don Dillman et al. 2010. "Research synthesis AAPOR report on online panels." *Public Opinion Quarterly* 74, 711-781.

Baker, Reg, J. Michael Brick, Nancy A. Bates, Mike Battaglia, Mick P. Couper, Jill A. Dever, Krista J. Gile, and Roger Tourangeau. 2013. "Summary report of the AAPOR task force on non-probability sampling." *Journal of Survey Statistics and Methodology*, smt008.

Bethlehem, Jelke. 2010. "Selection bias in web surveys." *International Statistical Review,* 78, 161-188.

Brick J.Michael. 2011. "The future of survey sampling". *Public Opinion Quarterly*, 75, 872-88.

Brick, J.Michael. 2015. "Compositional model inference". *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pp. 299-307. Alexandria, VA: American Statistical Association.

Buelens, B., Burger, J. and van den Brakel, J.A. 2018. "Comparing inference methods for non-probability samples." *International Statistical Review,* 10.13094/SMIF-2018-00003.

Callegaro, Mario, Reg Baker, Jelke Bethlehem, Anja S. Göritz, Jon A. Krosnick, and Paul J. Lavrakas. 2014. "Online panel research." *Online panel research: A data quality perspective*, 1-22.

Craig, Benjamin M., Ron D. Hays, A. Simon Pickard, David Cella, Dennis A. Revicki, and Bryce B. Reeve. 2013. "Comparison of US panel vendors for online surveys." *Journal of Medical Internet Research* 15, 11.

Crain, R., Cooper, C. and Dickinson, J.L. 2014. "Citizen science: a tool for integrating studies of human and natural systems." *Annual Review of Environment and Resources*, *39*, 641-645.

Dillman, D.A. 2017. "The promise and challenge of pushing respondents to the Web in mixed-mode surveys." *Survey Methodology*,*43*, 3-30.

Ding, Helen, Carla L. Black, Sarah W. Ball, Sara M.A. Donahue, Rebecca Fink, Carolyn B. Bridges, Katherine E. Kahn, Peng-Jun Lu, Walter W. Williams, Denise J. Jamieson. (2015). "Pregnant women and flu vaccination, Internet Panel Survey, United States, November 2015." Downladed from www.cdc.gov/flu/fluvaxview/pregnant-women-nov2015.htm on December 18, 2015.

Elliott, M.R. and Valliant, R. 2017. "Inference for nonprobability samples." *Statistical Science*, *32*, 249-264.

Fahimi, Mansour. 2015. "Towards a new math for sample surveys in the 21st Century." Paper presented at American Association of Public Opinion Research conference.

Groves, R.M., Couper, M.P., Presser, S., Singer, E., Tourangeau, R., Acosta, G.P. and Nelson, L. 2006. "Experiments in producing nonresponse bias." *International Journal of Public Opinion Quarterly*, *70*, 720-736.

Hamari, J., Koivisto, J. and Sarsa, H. 2014. "Does gamification work?--a literature review of empirical studies on gamification." In *System Sciences (HICSS), 2014 47th Hawaii International*

Conference, downloaded from https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6758978 on June 26, 2018.

Jiorle, R.P., Ahrens, R.N. and Allen, M.S. 2016. "Assessing the utility of a smartphone app for recreational fishery catch data." *Fisheries*, *41*, 758-766.

Knaub, James R. 2007. "Model and Survey Performance Measurement by the RSE and RSESP." *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 2730-2736.

Keusch, F. and Zhang, C. 2017. "A review of issues in gamified surveys." *Social Science Computer Review*, *35*, 147-166.

Lazer D, Kennedy R, King G, Vespignani A. 2014. "The parable of Google Flu: traps in big data analysis." *Science*, 343, 1203-1205.

Lee, Sunghee, and Richard Valliant. 2009. "Estimation for volunteer panel web surveys using propensity score adjustment and calibration adjustment." *Sociological Methods & Research* 37, 319-343.

Liu, B., Stokes, L., Topping, T. and Stunz, G. 2017. "Estimation of a total from a population of unknown size and application to estimating recreational red snapper catch in Texas." *Journal of Survey Statistics and Methodology*, *5*, 350-371.

Medway, R.L. and Fulton, J. 2012. "When more gets you less: a meta-analysis of the effect of concurrent web options on mail survey response rates." *Public Opinion Quarterly*, *76*, 733-746.

Mercer, A.W., Kreuter, F., Keeter, S. and Stuart, E.A. 2017. "Theory and practice in nonprobability surveys: parallels between causal inference and survey inference." *Public Opinion Quarterly*, *81*, 250-271.

Messer, B.L. and Dillman, D.A. 2011. "Surveying the general public over the internet using address-based sampling and mail contact procedures." *Public Opinion Quarterly*, *75*, 429-457.

Miller, Peter G., Jennifer Johnston, Matthew Dunn, Craig L. Fry, and Louisa Degenhardt. (2010). "Comparing probability and non-probability sampling methods in ecstasy research: Implications for the Internet as a research tool." *Substance use & misuse* 45, 437-450.

The National Academies of Sciences, Engineering, and Medicine. 2017. *Review of the Marine Recreational Information Program*. Washington, DC: The National Academies Press. https://doi.org/10.17226/24640.

NRC (National Research Council). 2006. *Review of recreational fFisheries survey methods*. Washington, DC: National Academies Press.

Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S. and Crowston, K. 2012. "The future of citizen science: emerging technologies and shifting paradigms." *Frontiers in Ecology and the Environment*, *10*, 298-304.

Neyman, Jerzy. 1934. "On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection." *Journal of the Royal Statistical Society*, 558-625.

Papenfuss, J.T., Phelps, N., Fulton, D. and Venturelli, P.A. 2015. "Smartphones reveal angler behavior: a case study of a popular mobile fishing application in Alberta, Canada." *Fisheries*, *40*, pp.318-327.

Rivers, D. 2007. "Sampling for web surveys". White paper prepared from presentation given at the *2007 Joint Statistical Meetings*, Salt Lake City, UT. https://s3.amazonaws.com/yg-public/Scientific/Sample+Matching_JSM.pdf.

Särndal, Carl-Erik, and Sixten Lundström. *Estimation in surveys with nonresponse*. John Wiley & Sons, 2005.

Särndal, C. E., B. Swensson, and J. Wretman. 1992. *Model assisted survey sampling*. Springer, New York.

Stephan, Frederick F., and Philip J. McCarthy. 1958. "Sampling opinions: An analysis of survey procedure." John Wiley & Sons.

Tourangeau R, Maitland A, Rivero G, Sun H, Williams D, Yan T. 2017. "Web surveys by smartphone and tablets: Effects on survey responses." *Public Opinion Quarterly*, 81, 896-929.

Valliant, Richard. 2013. "Comment." *Journal of Survey Statistics and Methodology*, smt010.

Venturelli, P.A., Hyder, K. and Skov, C. 2017. "Angler apps as a source of recreational fisheries data: opportunities, challenges and proposed standards." *Fish and Fisheries*, *18*(3), pp.578-595.

Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. 2014. "Forecasting elections with non-representative polls." *International Journal of Forecasting*.

Yeager, David S., Jon A. Krosnick, LinChiat Chang, Harold S. Javitz, Matthew S. Levendusky, Alberto Simpser, and Rui Wang. 2011. "Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples." *Public Opinion Quarterly*, nfr020.

Appendix

Marine Recreational Information Program

Fishing Effort Survey

Experimental Testing

9/26/2013

The MRIP Fishing Effort Survey (MFES) was implemented in Massachusetts, New York, North Carolina and Florida in October, 2012 to test a revised data collection design for monitoring marine recreational fishing effort. The survey, which collects information for two-month reference waves, included two experiments during the first two study waves, wave 5 (Sept-Oct 2012) and wave 6 (Nov-Dec, 2012), to test different survey design features aimed at maximizing efficiency and minimizing nonresponse error. Specifically, the experiments tested two versions of the survey instrument and four levels of cash incentives. Details of the experiments are provided below.

Instrument Testing

The MFES included an experiment to test two versions of the survey instrument. The objective of the experiment was to identify the instrument that maximized overall response rates while minimizing the potential for nonresponse bias resulting from differential nonresponse between anglers and non-anglers. One version of the instrument (Saltwater Fishing Survey) utilized a “screen out” approach that quickly identifies anglers (and non-anglers) and encourages participation by minimizing the number of survey questions, particularly for non-anglers. Person-level information, including details about recent fishing activity and limited demographic information, is collected for all household residents, but only if someone in the household reported fishing during the reference wave. The second version (Weather and Outdoor Activity Survey) utilized an “engaging” approach that encourages response by broadening the scope of the questions to include both fishing and non-fishing questions. This version collects person-level information for all residents of sampled households, regardless of whether or not household residents participated in saltwater fishing. Each wave, sampled addresses were randomly assigned to one of the two questionnaire types, which were evaluated in terms of response rates and reported fishing activity.

Table 1 provides the weighted response rates (AAPOR RR1 after excluding undeliverable addresses) and estimated fishing prevalence (percentage of households with residents who reported fishing during the wave) for the two versions of the instrument. Overall, the Weather and Outdoor Activity Survey achieved a significantly higher response rate than the Saltwater Fishing Survey, and there was no significant difference between instruments in estimated prevalence. The lack of a significant difference between instruments for estimated prevalence suggests that the gain in response for the engaging instrument cannot be attributed to increased survey participation by either anglers or non-anglers, but that both groups are more likely to respond to the Weather and Outdoor Activity Survey than the Saltwater Fishing Survey.

We also compared response rates and prevalence between instruments both among and within subpopulations defined by whether or not sampled addresses could be matched to state databases of licensed saltwater anglers – subpopulations expected to distinguish between households with anglers and households with no anglers or less avid anglers. As expected, both response rates and estimated prevalence were higher in the matched subpopulation than the unmatched subpopulation, confirming that a population expected to be interested in the survey topic - households with licensed anglers - is more likely to respond to a fishing survey and report fishing activity than a population that excludes licensed anglers[5]. Because we can identify household license status prior to data collection, we can account for differential nonresponse between matched and unmatched households in the estimation design by treating matched an unmatched domains as strata (Lohr, 2009).

Table 1. Weighted response rates and estimated prevalence overall and by domain for two versions of the survey instrument.

	Saltwater Fishing Survey		Weather and Outdoor Activity Survey	
	(%)	(n)	(%)	(n)
Response Rate				
Overall	31.1 (0.4)	17,511	34.7 (0.4)*	17,510
Matched	45.4 (1.1)	3,160	45.0 (1.0)	3,247
Unmatched	30.3 (0.4)	14,351	34.0 (0.5)*	14,263
Prevalence				
Overall	13.4 (0.5)	5,943	14.1 (0.5)	6,498
Matched	49.9 (1.7)	1,491	48.5 (1.6)	1,552
Unmatched	11.2 (0.6)	4,452	12.2 (0.6)	4,946

Notes – (1) standard errors are in parentheses. (2) Domains are defined by matching ABS samples to state databases of licensed saltwater anglers.
*Significantly different from Saltwater Fishing Survey ($p<0.05$).

There were no significant differences between instruments for either response rate or prevalence within the matched domain, suggesting that the inclusion of non-fishing questions in the Weather and Outdoor Activity Survey did not have an impact on response by either anglers or non-anglers. In the unmatched domain, the response rate was significantly higher for the Weather and Outdoor Activity Survey than the Saltwater Fishing Survey. However, the higher response rate did not translate to lower or higher estimates of prevalence; estimates of prevalence were not significantly different between instruments within the domain. This suggests that the engaging instrument uniformly increased the probability of response for anglers and non-anglers within the unmatched domain.

Differential nonresponse to a survey request between subpopulations will result in nonresponse bias if the subpopulations are different with respect to the survey topic. In the MRIP Fishing Effort Survey, we account for differential nonresponse between matched and unmatched households during sampling – matched and unmatched subpopulations are treated as independent

[5] The classification of sample into domains is dependent upon matching ABS sample to license databases by address and telephone number. This process is unlikely to be 100% accurate, so the unmatched domain is likely to include some households with licensed anglers. The unmatched domain also includes households with residents who fish without a license.

strata. Subsequently, the potential for nonresponse bias is limited to differential nonresponse between anglers and non-anglers within the matched and unmatched subpopulations. While the Weather and Outdoor Activity Survey achieved a higher response rate than the Saltwater Fishing Survey, both overall and within the unmatched subpopulation, the gains in response do not appear to result from a higher propensity to respond to the survey by either anglers or non-anglers. As a result, we cannot conclude that one of the instruments is more or less likely to minimize differential nonresponse between anglers and non-anglers. However, higher response rates decrease the risk for nonresponse bias and either lower data collection costs (for a fixed sample size) or increase the precision of estimates (for a fixed cost)[6]. Consequently, we conclude that the Weather and Outdoor Activity Survey is superior to the Saltwater Fishing Survey and recommend that the instrument be utilized for subsequent survey waves. Because it collects person-level information for all residents of all sampled households, the Weather and Outdoor Activity Survey also supports post-stratification of survey weights to population controls, which is an additional benefit of this recommendation.

Incentive Testing

The MRIP Fishing Effort Survey included an experiment to test the impact of modest, prepaid cash incentives on survey response and survey measures. Each wave, sampled addresses were randomly allocated to incentive treatment groups of $1, $2, and $5, as well as a non-incentive control group. Incentives were only included in the initial survey mailing. As in the instrument experiment, the objective of the incentive testing was to identify an optimum level of incentive that maximizes overall response while controlling costs and minimizes the potential for nonresponse bias resulting from differential nonresponse between anglers and non-anglers. Response rates, estimated fishing prevalence and relative costs of completing an interview were compared among incentive treatments to quantify the impacts of incentives.

Table 2 shows weighted response rates and the results of a logistic regression model predicting the effects of incentives on the odds of obtaining a completed survey. Including an incentive in the initial survey mailing significantly increased the odds of receiving a completed survey, and the odds increased significantly as the incentive amount increased. Cash incentives of $1, $2, and $5 increased the odds of receiving a completed survey by 63%, 93% and 137%, respectively.

Table 2. Weighted response rates and odds of receiving a completed survey by incentive amount.

Incentive	Response Rate (%)	n	Odds Ratio	95 % CI
$0	22.6	8,760	1.00	
$1	32.2	8,737	1.63*	(1.51, 1.77)
$2	36	8,738	1.93*	(1.78, 2.09)
$5	40.8	8,786	2.37*	(2.18, 2.56)

*Significantly different from the $0 control (*p*<0.05). Results of pairwise comparisons are as follows: $1>$0 (*p*<0.05), $2>$1 (*p*<0.05), $5>$2 (*p*<0.05).

[6] Assuming that fixed costs are the same for the two instruments, which was the case in the experiment.

Previous studies (Groves et al., 2006) have demonstrated that prepaid cash incentives can motivate individuals with little or no interest in a survey topic to respond to a survey request. Subsequently, we hypothesized that incentives would have a larger impact on non-anglers than anglers, minimizing differential nonresponse between the two populations. We initially explored this hypothesis by comparing estimated fishing prevalence among incentive conditions, expecting that gains in response in the incentive conditions would translate to lower estimates of fishing prevalence. The results do not support this hypothesis; there were no significant differences in prevalence among incentive conditions (Table 3).

Table 3. Overall estimated fishing prevalence by incentive amount.

Incentive	Prevalence (%)	n
$0	12.8	2,154
$1	14.1	3,065
$2	13.6	3,415
$5	14.1	3,807

Note – Differences in prevalence among treatments are not significant (p=0.05)

We further explored the interaction of topic salience and incentives by examining response rates and estimated fishing prevalence for the incentive conditions within domains defined by whether or not sampled addresses could be matched to databases of licensed saltwater anglers. We expected incentives to have a more pronounced effect in the unmatched domain, a population less likely to have an interest in the survey topic, than in the matched domain. Table 4 shows that incentives increased the odds of receiving a completed survey in both the matched and unmatched subpopulations. However, the value of the incentive seems to be more important in the unmatched domain, where the odds of receiving a completed survey increased uniformly and significantly as the value of the incentive increased ($0<$1<$2<$5). In contrast, the incentive amount was less significant in the matched domain, where the odds of receiving a completed survey were relatively flat among incentive conditions. These results are consistent with our expectations and suggest that a population with a low propensity to respond to a fishing survey can be motivated to participate by cash incentives, and that the motivation may increase as the incentive amount increases.

Table 4. Odds of receiving a completed survey by level of incentive for sample that could and could not be matched to state databases of licensed anglers.

	Subpopulation	
Comparison Pair	Matched OR	Unmatched OR
$1 vs. $0	1.75**	1.63**
$2 vs. $0	2.01**	1.93**
$5 vs. $0	2.11**	2.39**
$2 vs. $1	1.15	1.18**
$5 vs. $1	1.21*	1.46**
$5 vs. $2	1.05	1.24**

Notes – The second value in the comparison pair is the reference value.
Significance: *p<0.05, **p<0.0001

As noted previously, we expected that the gains in response in the incentive conditions would translate to lower estimates of fishing prevalence, particularly in the unmatched subpopulation. Once again, the results are not consistent with expectations; differences in fishing prevalence among treatments were not significant in either the matched or unmatched domain (Table 5). The lack of an effect of incentives on fishing prevalence suggests that the gains in response associated with increasing incentive amounts are uniform between anglers and non-anglers.

Table 5. Estimated fishing prevalence by incentive amount for a population of anglers (matched) and non-anglers (unmatched).

	Subpopulation			
	Matched		Unmatched	
Incentive	(%)	(n)	(%)	(n)
$0	49.2	533	10.7	1,621
$1	50.3	779	12	2,286
$2	48.6	837	11.6	2,578
$5	48.2	894	12.4	2,913

Note – Within subpopulations differences in prevalence among treatments are not significant (*p*=0.05)

We also examined the effect of cash incentives on overall data collection costs, specifically the direct costs of printing, postage, and the cash incentives themselves. Table 6 shows that the $5 incentive provided the largest gain in response, but the gain came at a relative cost of approximately $0.15 per completed interview. In contrast, the additional costs of the $1 and $2 incentives (20% and 38% higher cost than the $0 control, respectively) are more than offset by the associated gains in the number of completed surveys (42% and 58%, respectively). In other words, including a $1 or $2 cash incentive in the initial survey mailing actually decreased the cost of receiving a completed survey by 22% and 20%, respectively. These cost savings, which are conservative[7], could be used to lower overall data collection costs (for a fixed sample size) or increase the precision of survey estimates (for a fixed cost).

Table 6. Effect of incentives on data collection costs

Incentive Amount	Relative Cost Difference	Relative Difference in Completed Surveys	Relative Cost per Completed Survey
$0	1.00	1	$1.00
$1	1.20	1.42	$0.78
$2	1.38	1.58	$0.80
$5	1.90	1.75	$1.15

Note – relative differences reflect the ratio of quantities (cost, completes) in the experimental treatments to the zero dollar control.

Including a modest prepaid cash incentive in survey mailings clearly has a positive effect on survey response rates; the odds of receiving a completed survey increased significantly as the incentive amount increased. We expected the incentives to have a greater effect on non-anglers than anglers and decrease the potential for nonresponse bias by minimizing differential nonresponse between these two populations. However, the results of the experiment suggest that incentives increase response propensities for non-anglers and anglers equally. While this result does not support our hypothesis, it does demonstrate that incentives can increase the quantity of data without having a negative impact on survey measures. The experiment also demonstrated that incentives can decrease overall data collection costs. Based upon these findings, we conclude that a $2 incentive is optimal in terms of both maximizing response rates and minimizing data collection costs.

[7] The cost comparison assumes that the non-incentive direct costs (postage and printing) are the same for all survey treatments and does not reflect the fact that incentive conditions may not require as many follow-up mailings.

References

Groves, R.M., M.P. Couper, S. Presser, E. Singer, R. Tourangeau, G.P. Piani, N. Lindsay. 2006. Experiments in producing nonresponse bias. Public Opinion Quarterly, 70: 720-736.

Lohr, S. 2009. Multiple Frame Surveys. Chapter 4 in Pfeffermann, D. (Ed.) *Handbook of Statistics: Sample Surveys Design, Methods and Applications* (vol. 29A). Elsevier, Amsterdam.